1993-09

Application of logistic regression and survival analysis to the study of CEP, manpower performance and attrition

Lian, Tian Tse.

Monterey, California. Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

APPLICATION OF LOGISTIC REGRESSION AND SURVIVAL ANALYSIS TO THE STUDY OF CEP, MANPOWER PERFORMANCE AND ATTRITION

by

Lian Tian Tse

September, 1993

Thesis Advisor: S.Y. Sohn

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved OMB Np. 0704

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE September 1993	3. REPORT TYPE AND DATES COVERED Master's Thesis (Final)

4. TITLE AND SUBTITLE APPLICATION OF LOGISTIC REGRESSION AND SURVIVAL ANALYSIS TO THE STUDY OF CEP, MANPOWER PERFORMANCE AND ATTRITION	5. FUNDING NUMBERS
6. AUTHOR(S) Lian Tian Tse	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Ministry of Defence Singapore	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

This thesis is an application of logistic regression and survival analysis techniques to the study of current estimated potential (CEP), manpower performance, and attrition behaviour in the Singapore military. The manpower data includes both active (30%) and reserve personnel (70%) who entered service as early as the late fifties to as recent as the year 1992. The covariates under consideration are education level, academic or overseas military training award, current rank, length of service, rank seniority, age, salary grade, previous year's annual performance grade and CEP estimates.

The study identifies the covariates that explain the CEP and annual performance for the binary and polytomous models of the officers who were still on active duty as of 31 Dec 1992. It also examines the trend of attrition behaviour of officers using data from both the active and reserve personnel.

The results of the study show that (1)higher education level does not necessary result in better performance grade although it seems to give an indication of higher CEP, (2)The higher the rank of an officer, the more likely it is for him to have a poorer performance grade than when he was in the previous rank, (3)Education level is a significant covariate of the survival functions, and (4)Engineering officers generally has a higher attrition rate than the other service support officers.

14. SUBJECT TERMS logistic regression, manpower performance, current estimated potential, survival analysis, military attrition	15. NUMBER OF PAGES 109
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std 239-18

Application of Logistic Regression and Survival Analysis to the Study of CEP, Manpower Performance and Attrition

by

Lian Tian Tse
Ministry of Defence, Singapore
BSinME.,National University of Singapore, 1988

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1993

Author: [redacted]

Lian Tian Tse

Approved by: [redacted]

S.Y. Sohn, Thesis Advisor

[redacted]

G.F. Lindsay, Second Reader

[redacted]

Peter Purdue, Chairman
Department of Operations Analysis

ABSTRACT

This thesis is an application of logistic regression and survival analysis techniques to the study of current estimated potential (CEP), manpower performance, and attrition behaviour in the Singapore military. The manpower data includes both active (30%) and reserve personnel (70%) who entered service as early as the late fifties to as recent as the year 1992. The covariates under consideration are education level, academic or overseas military training award, current rank, length of service, rank seniority, age, salary grade, previous year's annual performance grade and CEP estimates.

The study identifies the covariates that explain the CEP and annual performance for the binary and polytomous models of the officers who were still on active duty as of 31 Dec 1992. It also examines the trend of attrition behaviour of officers using data from both the active and reserve personnel.

The results of the study show that (1)higher education level does not necessary result in better performance grade although it seems to give an indication of higher CEP, (2)The higher the rank of an officer, the more likely it is for him to have a poorer performance grade than when he was in the previous rank, (3)Education level is a significant covariate of the survival functions, and (4)Engineering officers generally has a higher attrition rate than the other service support officers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Manpower planners and recruitment agencies in the Singapore's DOD are keen to identify the various explanatory variables that could be used to explain current estimated potential (CEP - an officer's estimate of his command capacity by 45 years of age) and performance. The Government's advocacy for family planning in the seventies has resulted in a reduction of eligible males who could be recruited for a military career in the nineties. If the current attrition of military officers is not properly checked, then at the turn of the century the military would have a mammoth task in keeping up with its operational manning requirements. Identifying the significant covariates and trends of attrition would greatly assist the responsible agencies in force planning and formulation of manpower policies.

In view of the above, two techniques are employed in this thesis. First, the logistic regression technique is used to identify the significant covariates that could explain and predict CEP and performance grade. Two models are considered, namely, the binary and polytomous logistic regression. The covariates under consideration are education level, academic or overseas military training award, current rank, length of service, rank seniority, age, salary grade, previous year's annual performance grade and CEP estimates.

Second, the survival analysis technique is used to analyze the trend of the attrition behaviour of officers who entered service during the period 1965-70, 1971-76, and 1977-82. The graphical approach is used to examine the attrition trends which does not

require any statistics background. However, formal statistical tests are conducted to ascertain the visual observations.

For the CEP binary response model, the data is divided into two groups. The first group consists of officers who have a CEP rank of Major and below while the second group consists of officers who have a CEP rank of at least a Lieutenant Colonel. A standard measure of the quality of model prediction using a cutoff point of *0.40* resulted in approximately *87%* correct classification for each group. As for the performance binary response model, the data is also divided into two groups. The first group consists of officers who have a performance grade of B minus and below in the 1992 performance appraisal. The second group consists of officers who have a performance grade of at least a B in the 1992 performance appraisal. A cutoff point of *0.64* would result in each group being approximately *74%* correctly classified.

The CEP polytomous response model has an *82%* correct prediction capability when the fitted model is tested on a second population of officers as compared to *68%* for the performance model.

The significant findings are outlined below.

- **Education Level-** Education level is not a significant predictor of performance though a higher education level seems to give an indication of higher CEP.
- **Training Award-** There is insufficient evidence to support the notion that officers given an academic or overseas military training award tends to have a better performance grade than those who did not receive any.
- **Rank-** The higher the rank of an officer, the more likely it is for him to get a poorer performance grade than when he was in the previous rank.

- **Previous year's CEP and Performance Grade-** Current year's CEP estimation and performance grade prediction are highly correlated to previous year's CEP and performance grade.

The results of the survival analysis are briefly outlined below.

- **Non-Graduate vs Graduate-** The attrition behaviour in each of the three enlistment periods (officers who entered service during 1965-1970, 1971-1976, and 1977-1982) between non-graduates and graduates is not significantly different.

- **Education Level-** Education level has a strong relationship with the attrition behaviour of the officers. Officers with a Cambridge General Certificate of Education (GCE) 'O-' or 'A-' level qualification have consistently survived longer in the service than officers who have other educational qualifications. On the contrary, officers with diploma qualification exhibit the lowest survival functions.

- **Training Award-** The trend of the difference in the survival functions between non-award and award holders for the three enlistment periods is statistically the same.

- **Support Vocation-** The Engineering and Air Force support officers have the highest attrition rate during the first year of service. It drops to the lowest at the beginning of the third year, after which the attrition rates of the Engineering officers are generally higher than the other two categories of officers. The Army support officers exhibit a relatively constant attrition rate throughout the entire period of study.

- **Service Group-** For the first six years of service, the Naval officers have a lower risk of leaving the service than their Army counterparts. In contrast, after the first six years, the converse is true.

I. INTRODUCTION

A. BACKGROUND

In manpower studies, much attention is given to job changes, layoffs, retirements, performance appraisal, and promotions. Very often performance appraisal and promotion go hand-in-hand. In Singapore's military organization, staff performance appraisal is carried out annually. The military officers' promotions are based on this annual assessment.

The annual assessment consists of two parts. The first part assesses the officer's aggregated annual performance appraisal which encompasses job performance, work attitudes and personal qualities. Job performance is being assessed through factors such as initiative, planning ability, applied knowledge, quality of work, and decision making. Work attitude is being assessed through factors like drive and determination, responsibility, and teamwork. Personal qualities is being assessed through factors like the officer's writing ability, oral expression, stability in stressful situations, human relations, and last but not least leadership qualities. All these factors are given on a numeric scale with 1 being the highest possible and 7 being the lowest. All that is required of the reporting officer is to tick the box corresponding to the score to be awarded to that particular factor under consideration. Finally, the overall performance is an aggregate score based on the assessment of job performance, work attitudes, and

personal qualities. It is given on a numeric scale from 1 to 15 with 15 representing an A^+, 14 an A, 13 an A^-, 12 a B^+, 11 a B, and so on.

The second part assesses the officer's current estimated potential (CEP). The CEP measure is a military rank assessment. It is an estimate of the officer's command capacity by 45 years of age (e.g. Rank: LTC, Appointment: Bn Comd/CO of Trg School). This assessment is independent of the above performance appraisal. Here, the officer is being assessed on his ability to approach a problem from a higher vantage point (known as the Helicopter Quality). This includes his ability to detect quickly and attend to relevant details within a broader context, and be constantly able to provide solutions of good vision. The officer is also assessed on his powers of analysis, imagination, and sense of reality when faced with complex and unfamiliar problems.

B. PROBLEM STATEMENT

Currently, not much work has been done in the area of annual CEP and performance prediction of combat officers. Manpower planners and recruitment agencies are keen to identify the various explanatory variables that could be used to explain CEP and performance. In many military organizations, education level has by far proved to be a valuable predictor of performance. Is education level a valuable predictor of performance in the Singapore context? Is education level also a good explanatory variable for CEP estimation?

Some of the officers are awarded academic or overseas military training to increase their knowledge and professionalism during their careers as military officers.

Will an officer who is given such an award perform significantly better than those that are not given any training award?

Another area of interest asks whether there is any significant difference in performance and CEP among officers of differing vocations.

Family planning in the seventies has drastically reduced the population of eligible males which could be recruited for a military career in the nineties. The military has to compete with the civilian organizations for this limited pool of resource. To alleviate the problem of manpower shortages, the military has to ensure that the attrition level of the officers is under control. A high attrition level will disrupt the efficiency and readiness of the military as a whole. It is also costly since new officers have to be recruited and time is needed to train them to a proficiency level compatible to their predecessors. Hence, the factors that affect the length of service of an officer is also of great interest to the military commanders, manpower planners and recruitment agencies. Identifying these factors could greatly assist the responsible agencies in force planning and formulation of manpower policies.

C. THESIS OVERVIEW

1. Objective

This thesis examines the relationship between an officer's covariates (past performance and CEP assessments, education level, training award, current rank, seniority in current rank, age, length of service) and (a) future CEP estimation, and (b) the prediction of future performance. It also investigates the attrition behaviour of

officers who entered service during the period from 1965-70, 1971-76, and 1977-1982 as a function of educational level, training award, support vocation and service type in general.

The primary interest is to identify those covariates that could significantly explain an officer's CEP assessment and performance appraisal. The secondary interest is to examine the attrition pattern of officers who entered service during the periods from 1965-70, 1971-76, and 1977-82.

2. Methodology

The study is basically divided into two parts. The first part uses the logistic regression technique to estimate CEP and predict performance. The simplest model is the binary response model. It is used to model dichotomous outcomes, as for example, whether an officer's CEP estimate would be of Major (MAJ) rank and below, or Lieutenant Colonel (LTC) rank and above. In contrast, the polytomous response model is able to provide us with more information. The response is no longer restricted to two levels. In this thesis, the CEP model has four levels namely, CPT, MAJ, LTC, and, COL and above. The tradeoff for the polytomous response model is that the model is difficult to evaluate and explain to the novice.

The second part of the study uses survival analysis techniques to compare the attrition patterns of officers who are enlisted in the three different periods. This thesis examines only the individual effects of each covariate, namely, education, training award, support vocation, and service group.

3. Findings

The binary and polytomous response models both give the same significant covariates. For the CEP model, the significant covariates are education, current rank, rank seniority, age, and previous year's annual performance grade and CEP. For the performance model, the significant covariates are current rank, rank seniority, and previous year's annual performance grade and CEP.

For the CEP response model, the findings indicate that it is more likely for a highly educated, young high-ranking officer to have a CEP estimate of at least a LTC rank. Additionally, the higher the previous year's performance grade and CEP estimate, the higher the probability that the officer's CEP is at least a LTC rank.

For the annual performance response model, an interesting result is found. The higher the rank of an officer, the more likely it is for him to have a poorer performance grade than when he was in the previous rank. This could be a direct result of quotas placed on the performance grades.

Education level is found to have a significant effect on the attrition behaviour of the officers for the three enlistment groups under study. Generally, the Engineering Support officers seems to have a higher risk of leaving the service than the Army and Air Force Support officers.

4. Organization

The organization of this thesis follows the order in which the study was performed. Chapter II describes the methodology of binary and polytomous logistic regression, and the survival analysis technique used in the thesis. Chapter III gives a

summary of the exploratory analysis of the population under study. It also contains a brief description of the covariates and a code book. Chapter IV presents the binary models for future CEP estimation and performance prediction. Evaluation of the models developed are also discussed in details. Chapter V presents the polytomous models for future CEP estimation and performance prediction. Chapter VI contains analyses of single covariate effect on the attrition behaviour of officers enlisted during three different time periods. Chapter VII contains the conclusions and a summary of the findings, together with the recommendations for future work.

II. METHODOLOGY

A. LOGISTIC REGRESSION

Linear logistic regression is one of the many special cases of generalized linear models. It is characterized by three components: a random component, which identifies the probability distribution of the response variable; a systematic component, which specifies a linear function of explanatory variables that is used as a predictor; and a link function describing the functional relationship between the systematic component and the expected value of the random component. [Ref. 1:p. 80]

Linear logistic regression technique fits the model for binary or ordinal response data using the method of maximum likelihood. Logistic regression model has been in use in statistical analyses for many years. It is frequently used when an individual is to be classified into two or more groups. In the past, logistic regression found most of its application in the medical field [Ref. 2:p. vii]. It has been used, for example, to predict the survival of critically ill patients who are admitted to an intensive care unit as a function of certain physiological variables. Its application has expanded from health sciences to many other fields such as sociology, criminology, marketing and manpower studies.

The fundamental assumption in linear logistic regression analysis is that natural logarithms of odds is linearly related to the independent covariates. Here, odds is

defined as the ratio of the probability of an event occurrence to the probability of non-occurrence of the event.

Variable selection is necessary when there are many candidate covariates for model building. Three commonly used methods are: forward selection, backward elimination, and stepwise selection. In this thesis, the stepwise variable selection procedure of the Statistical Analysis System (SAS) software package is used for variable selection. The stepwise method combines both the forward selection and backward elimination methods. [Ref. 3:p. 196]

1. Binary Response Model

In the binary response model, the response variable is binary or dichotomous. An individual can take on one of the two possible values, denoted for convenience by 0 and 1. Observations of this nature arise, for instance, an individual has either been promoted (Y=1) or has not (Y=0) in the annual staff promotion exercise. We may then define

$$pr(Y=0) = 1 - \pi; \qquad pr(Y=1) = \pi \tag{1}$$

for the probabilities of 'failure' (not promoted) and 'success' (promoted) respectively. The probability of an officer's promotion would be related to his characteristics such as annual performance grade and CEP.

The goal of this analysis is to find the best fitting and most parsimonious yet practical and reasonable model to describe the relationship between the response variables (annual performance grade, and CEP) and a set of independent explanatory

variables. These independent variables are often known as covariates. The term "explanatory variable" will be used interchangeably with "covariate" throughout this thesis.

A wide choice of link functions $g(\pi)$ is available to describe the functional relationship between the probability distribution of the response variable and the linear function of explanatory variables [Ref. 4:p. 108]. Three functions commonly used in practice are:

- the logit or logistic function

$$g_1(\pi) = \log\{\pi/(1 - \pi)\};$$

- the probit or inverse Normal function

$$g_2(\pi) = \Phi^{-1}(\pi); \text{ and}$$

- the complementary log-log function

$$g_3(\pi) = \log\{-\log(1 - \pi)\}.$$

A fourth possibility, the log-log function

$$g_4(\pi) = -\log\{-\log(\pi)\},$$

which is the natural counterpart of the complementary log-log function, is seldom used because its behaviour is inappropriate for $\pi < ½$, the region that is usually of interest. All four functions can be obtained as the inverse of well-known cumulative distribution functions having support on the entire real axis. The first two functions are symmetrical in the sense that

$$g_1(\pi) = -g_1(1 - \pi).$$

The later two functions are not symmetrical in this sense, but are related via

$$g_3(\pi) = -g_4(1 - \pi).$$

a. Advantages of Logistic Function

The logistic function is used in this thesis because of its simple interpretation as the logarithm of the odds ratio, $\pi/(1 - \pi)$. Apart from this, the logistic function has one important advantage over all alternative transformations in that it is eminently suited for the analysis of data collected retrospectively. [Ref. 4:p. 109]

b. Parameter Interpretation

If a linear logistic model is used with p covariates, then we would have the model

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p \quad , \tag{2}$$

for the log odds of a positive response ('success' or say, promoted). Throughout this thesis, the term "log" refers to the "natural logarithm". Equivalently, in terms of the probability of belonging to a positive response, Equation (2) can be rewritten as

$$\pi=\frac{\exp(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p)}{1+\exp(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p)} \quad . \tag{3}$$

This is the inverse function of $g_1(\pi)$. Assuming that the covariates are functionally unrelated, the effect of a unit change in x_2 is to increase the log odds by an amount β_2. In other words, we may say that a unit change in x_2 has the effect of increasing the odds of a positive response multiplicatively by the factor $\exp(\beta_2)$. It is important that

all the other covariates (i.e. $x_1, x_3, \ldots, x_p$) are held fixed and not be permitted to vary as a consequence of the change in x_2. [Ref. 4:p. 110]

2. Polytomous Response Model

If the response of an individual or item is restricted to one of a fixed set of possible values, we say that the response is polytomous. The binary response model is a special case of the polytomous response model. In the development of models for polytomous response variable, we need to know its underlying measurement scale. Many methods are available for modelling nominal scaled response variable (performance grade) but will not be discussed here [Ref. 2:p. 216]. In this thesis, methods for modelling ordinal scaled response variable (CEP) is presented.

When response categories have a natural ordering, logit models should utilize that ordering. A familiar example of ordinal response category is the rating scales used in food testing and wine tasting.

a. Cumulative Logit Model - Proportional Odds Model

All the K-1 cumulative logits for a K-category response variable are incorporated into a single, parsimonious model. The simplest models in this class involve parallel regressions on the chosen scale, such as

$$\log\left(\frac{\gamma_j(x)}{(1-\gamma_j(x))}\right) = \theta_j - \beta^T x, \qquad j=1,\ldots,k-1\,, \qquad \textbf{(4)}$$

where $\gamma_j(x) = \mathrm{pr}(Y \le j \mid x)$ is the cumulative probability up to and including category j, when the covariate vector is x. The negative sign in (4) is a convention ensuring that

large values of $\beta^T x$ lead to an increase of probability in the higher numbered categories. Both θ and β in (4) are treated as unknown, and θ must satisfy $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_{k-1}$ [Ref. 4:p. 153]. Model (4) is known as the proportional-odds model because the ratio of the odds of the event $Y \leq j$ at $x = x_1$ and $x = x_2$ is

$$\frac{\gamma_j(x_1)/(1 - \gamma_j(x_1))}{\gamma_j(x_2)/(1 - \gamma_j(x_2))} = \exp(-\beta^T(x_1 - x_2)), \tag{5}$$

which is independent of the choice of category (j). The odds ratio of cumulative probabilities in (5) is called a cumulative odds ratio. The log of the cumulative odds ratio is proportional to the distance between the values of the explanatory variables, with the same proportionality constant applying to each cutpoint. Its interpretation is that the odds of making response $\leq j$ are $\exp[-\beta^T(x_1 - x_2)]$ times higher at $x = x_1$ than at $x = x_2$.

B. SURVIVAL ANALYSIS

Statistical methods for survival analysis have evolved largely from biomedical and epidemiologic studies of humans and animals. Survival analysis is often used to analyze data on the length of time it takes for a specific event to occur. Survival time can be broadly defined as the time to the occurrence of a given event of interest. This event can be the death of a person, animal, or insect; or the termination of employment.

Survival data may include subjects in the study who have not experienced the event of interest at the end of the study or time of analysis. For instance, some patients may still be alive at the end of a study period. For these subjects, the exact survival

times are unknown. These are called censored observations or censored times and can also occur when individuals are lost to follow-up, in that they fail to turn up for subsequent medical review after a period of study. It would be impractical to wait until every subject has died before conducting any analysis. This is an intrinsic characteristic of survival data.

The attrition behaviour of military officers is analogous to what was described in the previous paragraph. The survival time of an officer is the length of service time prior to leaving the service and becoming a reserve. The officers that are still active at the end of the study period are treated as censored observations.

1. Survival Functions

In this analysis, it is assumed that the survival time of an officer is discrete and represented by, t, (t=1,2,...,25), where t is the number of years of active service prior to going into reserve. The values of t are rounded to the next higher integer value. Therefore, if an officer went into reserve after serving 3.4 years of active duty, the survival time is 4 years.

If there are no censored observations, the survival function is estimated as the proportion of officers surviving longer than t and is given by

S(t) = P(an individual survives longer than t), where

$$S(t) = 1 - \left(\frac{\text{number of officers with surviving time} \leq t}{\text{total number of officers}}\right). \tag{6}$$

When censored observations are present, the numerator of (6) cannot always be determined. Nonparametric methods of estimating S(t) for censored data have to be used instead. [Ref. 5:p. 86]

2. Nonparametric Methods of Estimating Survival Functions

Many authors use the term life-table estimates for the product-limit (PL) estimates. The only difference is that the PL estimate is based on individual survival times while in the life-table method survival times are grouped into intervals. The PL estimate can be considered as a special case of the life-table estimate where each interval contains only one observation. It is more convenient to perform life table analysis when the data have already been grouped into intervals or the sample size is huge, say in the thousands.

The conditional proportion dying ($\hat{q}_i$) is defined as d_i/n_i for $i = 1,...,s-1$, and $\hat{q}_s = 1$, where d_i is the number of individuals who die in the ith interval and *n_i is the number of individuals who are exposed to risk in the ith interval. It is an estimate of the conditional probability of death in the ith interval given exposure to the risk of death in the ith interval. The estimate of cumulative proportion surviving (survival function) at t_i is given by

*The number of individuals entering the first interval $\acute{n}_i$ is the total sample size. For subsequent intervals, the number of individuals entering the ith interval is equal to the number of individuals studied at the beginning of the previous interval minus those who are lost to follow-up, are withdrawn alive, or have died in the previous interval.

$$\hat{S}(t_i) = \prod_{j=1}^{i} (1 - \hat{q}_j) \quad . \tag{7}$$

This estimate was derived by Kaplan and Meier (1958), and in practice is often referred to as the *Kaplan-Meier estimate*. [Ref 6]

3. Hazard Function

The hazard function for the *i*th interval, estimated at the midpoint, is

$$\hat{h}(t_{mi}) = \frac{d_i}{b_i(n_i - \frac{1}{2}d_i)} = \frac{2\hat{q}_i}{b_i(1 + \hat{p}_i)} \quad , \qquad i=1,\ldots,s-1 \; , \tag{8}$$

where

- b_i is the width of each interval,
- d_i is number of individuals who died in the *i*th interval,
- n_i is number of individuals who are exposed to risk in the *i*th interval,
- $\hat{p}_i$ is conditional proportion surviving and is defined as $\hat{p}_i = 1 - \hat{q}_i$, which is an estimate of the conditional probability of surviving in the *i*th interval, and
- $\hat{q}_i$ is conditional proportion dying and is defined as the ratio of d_i over n_i.

The above equation (8) is the number of deaths per unit time in the interval divided by the average number of survivors at the midpoint of the interval. The hazard function is also commonly known as the instantaneous failure rate. It is a measure of the risk of failure at a point in time during the aging process.

Before proceeding to analyze the data using the various techniques introduced earlier, some mention of the data set is desirable. This is taken care of in the following chapter.

III. DATA OVERVIEW

This chapter gives a brief description of the data set that consists of about 17000 records of individual officer's characteristics. This data set contains records of both Singapore's active and reserve officers for the period from 1959 to 1992.

A. POPULATION

The models for CEP estimation and performance prediction consider both the male and female officers who were still in active duty on 31 Dec 1992. Since the female population is relatively small compared to the male counterparts, the study does not discriminate between the two sexes. Out of the total of about 17000 records, about 30% of them are still active.

Table 1 shows the distribution of actual CEP of the active officers from 1990 to 1992. Table 2 shows the distribution of actual annual performance of the active officers for the same period.

From the two tables, it can be observed that the percentages of individuals in each response category over the three years are more or less the same. Additional two-way tables of CEP and performance as a function of educational level, award, age group, length of service, and rank seniority are found in Appendix A.

Table 1. CEP DISTRIBUTION FROM THE YEAR 1990 TO 1992

CEP (RANK)	PERCENT		
	1990	1991	1992
CPT	1.9	2.7	3.0
CPT^+	0.7	0.7	0.9
MAJ	29.6	28.8	25.2
MAJ^+	17.9	14.9	15.6
LTC	34.5	36.3	35.3
LTC^+	9.5	10.4	13.5
COL	4.9	5.0	4.9
COL^+	0.9	1.0	1.5
BG	0.1	0.2	0.1
MG	-	0.1	0.1

Table 2. PERFORMANCE DISTRIBUTION FROM THE YEAR 1990 TO 1992

PERFORMANCE APPRAISAL GRADE	FREQUENCY (PERCENT)		
	1990	1991	1992
E (2)	0.1	0.1	-
D (5)	4.4	6.0	4.9
C^- (7)	-	0.1	-
C (8)	44.6	43.8	42.8
C^+ (9)	16.1	15.4	16.1
B (11)	25.2	24.2	24.6
B^+(12)	7.4	8.6	9.7
A (14)	2.3	1.7	1.9

Note: The figures in brackets represent the numeric score given on the performance appraisal form.

B. COVARIATES

There are altogether eight covariates considered in this study. Except for length of service, rank seniority and age which are continuous variables, all the remaining covariates are categorical. Here is a brief description of the covariates:

- Education Level - The education level of the officers varies from the Cambridge General Certificate of Education (GCE) 'O' level to Doctorate. About 86% of the active officers have at least a GCE 'A' level or diploma qualification. Thirty-three percent of the active officers have at least a graduate degree.

- Academic or Overseas Military Training Awards - About 30% of the officers received some form of academic or overseas military training awards. Overseas military training awards include Sandhurst (United Kingdom), West Point (United States), the Naval Academy (United States), to name a few. Academic training awards include both local and overseas universities.

- Rank - 'Rank' is the rank of an officer as of 31 Dec 1992. It ranges from the rank of Lieutenant to the rank of Major General.

- Length of Service - The length of service (measured in years) is computed from the year an officer first enters the military service as a recruit to 1992.

- Rank Seniority - Rank seniority is the number of years an officer has been in his most recent rank since last promotion.

- Age - 'Age' is the age of the officer.

- Salary Grade - The salary grade ranges on an ascending scale of 1 to 10. A higher grade in each of the rank will mean higher renumeration for an officer.

C. CODE BOOK

A code book for the individual officer's characteristics is given in Table 3.

Table 3. CODE BOOK FOR INDIVIDUAL OFFICER'S CHARACTERISTICS

VARIABLE	UNITS	SCALE	COMMENTS
ID	none	nominal	Officers numbered sequentially
EDU	none	ordinal	0 = unknown; 1 = GCE 'O' or equiv. and below; 2 = GCE 'A' or equiv.; 3 = Diploma and Adv. Diploma; 4 = General Degree; 5 = Honors Degree; 6 = Masters Degree; 7 = Doctorate
AWARD	none	nominal	1 = no award; 2 = academic or military training award
LGSVC	years	ratio	Length of service as at 31 Dec 92
RSNR	years	ratio	Number of years in the rank held since last promotion
AGE	years	ratio	Age as at 31 Dec 1992
SGD	none	ordinal	Salary grade in ascending order from 1 to 10
C89 to C92	none	ordinal	Current Estimated Potential, 1989 to 1992; 1 = CPT; 2 = Snr. CPT; 3 = MAJ; 4 = Snr. MAJ; 5 = LTC; 6 = Snr. LTC; 7 = COL; 8 = Snr. COL; 9 = BG; 10 = Snr. BG; 11 = MG
P89 to P92	none	ordinal	Performance Appraisal, 1989 to 1992; 1 = E^-; 2 = E; 3 = E^+; 4 = D^-; 5 = D; 6 = D^+; 7 = C^-; 8 = C; 9 = C^+; 10 = B^-; 11 = B; 12 = B^+; 13 = A^-; 14 = A; 15 = A^+

The code book is used for cross-reference when one does not understand what the number(s) in the data set means. This is the most important document in the data preparation phase. Once the code book has been prepared we can proceed to analyze the data. The next two chapters analyze the data set using the Logistic Regression technique.

IV. BINARY RESPONSE MODEL

A. CURRENT ESTIMATED POTENTIAL

The primary goal is to determine the covariates that can best explain the variation of CEP of an officer. The stepwise regression technique is used for variable selection. The significance levels for entry and staying in the model are set at $\alpha = 0.10$ and 0.12 respectively.

The response variable is the CEP for the year 1992 (denoted by CEP92). A response value of zero (Y=0) means a CEP estimate of MAJ^+ and below while a response value of one (Y=1) means a CEP estimate of LTC and above. This classification is chosen because the population under study can be approximately divided equally into these two groups (see Table 1 on page 18). In the process of model building three sets of candidate covariate combinations will be thoroughly investigated. They are

- Education level, training award, rank, length of service, rank seniority, age, salary grade, CEP grades from 1989 to 1991, and performance grades from 1989 to 1991,
- Education level, training award, rank, length of service, rank seniority, age, salary grade, CEP for the year 1991, and performance grade for the year 1991, and
- Education level, training award, rank, length of service, rank seniority, age, and salary grade.

A comparison of the models derived from the above three covariate combinations is given in detail and is presented in Section A of Appendix B. In this analysis, an event occurs when an officer is classified as having a CEP estimate of MAJ⁺ and below and a non-event when the officer have a CEP estimate of LTC and above. For convenience, the MAJ⁺ and below group is designated by MAJ, and LTC and above group by LTC. This convention will be adopted throughout this thesis.

The probability of being classified as MAJ is estimated by

$$P_{MAJ} = \frac{\exp[6.396-0.19E-2.26R-0.21S+0.22A-0.18(P91)-1.45(C91)]}{1+\exp[6.396-0.19E-2.26R-0.21S+0.22A-0.18(P91)-1.45(C91)]}.$$

Conversely, the probability of being classified as LTC is estimated by

$$P_{LTC} = \frac{1}{1+\exp[6.396-0.19E-2.26R-0.21S+0.22A-0.18(P91)-1.45(C91)]},$$

where

- E is educational level,
- R is current rank as at 31 Dec 1992,
- S is number of years in current rank since last promotion,
- A is age (in years) as at 31 Dec 1992,
- P91 is performance appraisal for the year 1991, and
- C91 is current estimated potential for the year 1991.

A unit change in the educational level has the effect of increasing the odds of being classified as MAJ multiplicatively by a factor of 0.82. In other words, the higher the educational level of an officer, the more likely he or she would belong to LTC.

Similarly, the higher the rank, rank seniority, performance grade and CEP in the previous year, the higher the probability that an officer would belong to the LTC group. On the contrary, a unit increase in age has the effect of increasing the probability of an officer belonging to MAJ group.

B. PERFORMANCE

For this model, the response variable is the performance grade for the year 1992 (denoted by PERF92). A response value of zero (Y=0) means a performance grade of B minus and below while a response value of one (Y=1) means a performance grade of at least a B. Like the CEP model, the same three covariate combinations are investigated. Again, for convenience, a response value of zero is designated as Group I while a response value of one is designated as Group II.

Coincidentally, the model selected is again derived from the second covariate combination. A comparison of the models derived from the three covariate combinations are discussed in Section B of Appendix B.

The probability of being classified as Group I is estimated by

$$P_I = \frac{\exp[7.1631+1.1R-0.28S-0.44(P91)-0.79(C91)]}{1+\exp[7.1631+1.1R-0.28S-0.44(P91)-0.79(C91)]} .$$

Conversely, the probability of being classified as Group II is estimated by

$$P_{II} = \frac{1}{1+\exp[7.1631+1.1R-0.28S-0.44(P91)-0.79(C91)]} ,$$

where

- R is current rank as at 31 Dec 1992,
- S is number of years in current rank since last promotion,
- P91 is performance appraisal for the year 1991, and
- C91 is current estimated potential for the year 1991.

An interesting result is that a unit change in the rank of the officer to the next level will increase the odds of getting a performance grade of B minus and below multiplicatively by a factor of three. In other words, as an officer gets promoted to the next rank, the more likely his annual performance grade will deteriorate when compared with those in his previous rank. The remaining three covariates in the model, however, have the reverse effect.

C. EVALUATION OF THE MODEL

In a statistical model building analysis, it is in the interest of the investigator to know how much to trust the predictions derived from the model. The question commonly asked: Can the model predict correctly a high proportion of the time? Statistical significance does not necessarily mean that the model will predict very well since these measures are based on the model. Very often, results obtained that are statistically significant do not predict very well when implemented in the real world.

Equation (3) on Page 10 is the linear logistic model given in terms of the probability of belonging to a positive response (i.e., an event). In order to classify the officers into the two groups, a cutoff point must be determined, usually by graphical

means. This cutoff point is a probability ranging between 0 and 1, and is usually denoted by P_c. The cutoff point is chosen so that a high percentage of correct prediction is achieved for the two groups. An officer would be classified as MAJ group (for the performance model: performance grade of B minus and below) if the probability of an event is greater than or equal to P_c. The classification table in the SAS output (see Appendix C) provides information on sensitivity[*], specificity[+], false positive rate[**] and false negative rate[++].

1. Current Estimated Potential

Naturally, one would wish the percent correctly classified in each group to be as close to one as much as possible. Figure 1 gives the graphical representation of the prediction of percent correct plotted against the cutoff point. For example, for a cutoff point of about *0.40*, each group is approximately *87%* correctly classified. This may be a good choice of a cutoff point because it treats both groups equally. In contrast, a cutoff point of *0.04* would result in *99%* of the MAJ group classified correctly but only about *29%* of the LTC group.

The receiver operating characteristic (ROC) curve is a plot of the proportion of events (MAJ group) correctly classified as event (MAJ group) against the proportion of non-events (LTC group) incorrectly classified as event (MAJ group). Similarly, we

---

*Sensitivity is the proportion of event that were predicted to be event.

+Specificity is the proportion of non-event that were predicted to be non-event.

**False positive rate is the proportion of predicted event responses that were observed as non-event.

++False negative rate is the proportion of predicted non-event responses that were observed as event.

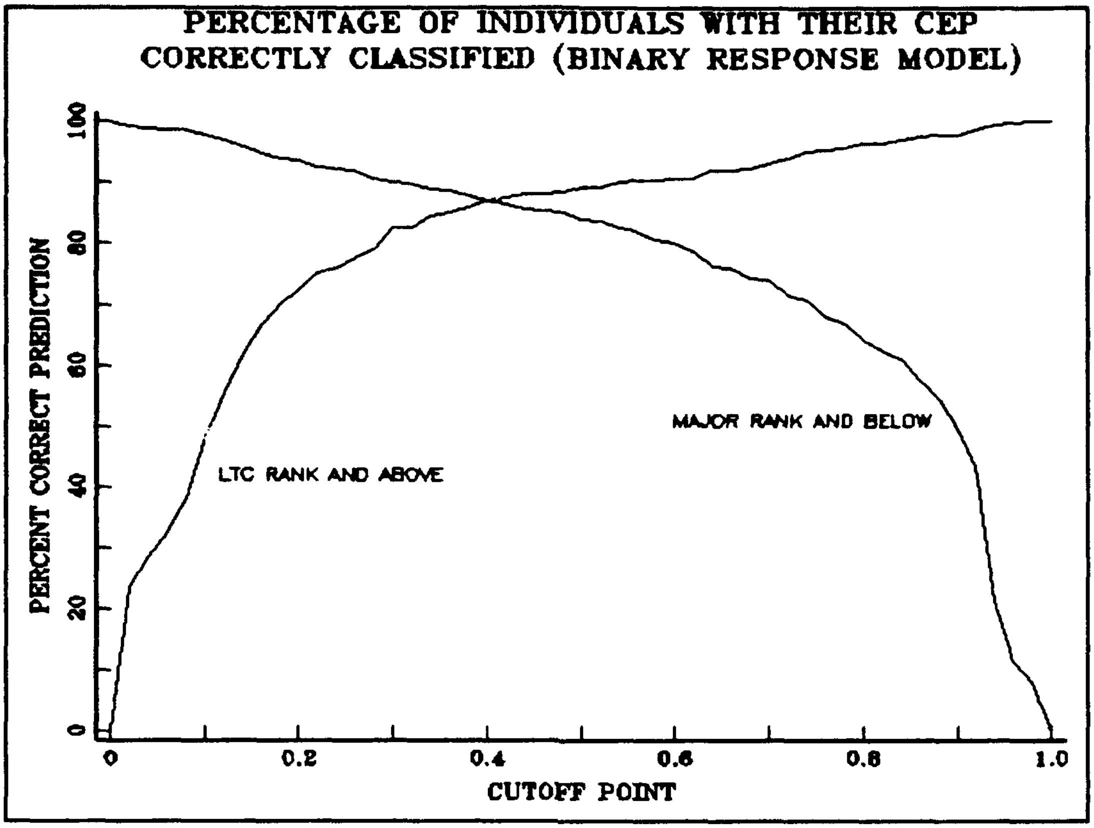

Figure 1: Percentage of Individuals with their CEP Correctly Classified (Binary Response Model).

could also plot the proportion of non-events correctly classified as non-event against the proportion of events incorrectly classified as non-events. Figure 2 gives these two ROC curves. In the top plot of Figure 2, the top curve represents the actual curve obtained from the prediction of an event based on the six variables obtained from the stepwise selection procedure (i.e., education level, rank, rank seniority, age, previous year performance grade and CEP estimate). The hypothetical curve (straight line) represents the chance-alone assignment (i.e., flipping of a fair coin). Likewise, the top curve of the bottom plot in Figure 2 represents the actual curve obtained from the prediction of an officer being classified as LTC.

From the plots in Figure 2, one can see that the model derived gives pretty good prediction. If a cutoff point of *0.4* is used, *87%* of both groups could be correctly classified with a false positive rate of *16%* and a false negative rate of *11%*. In other words 16% of the LTC group would be incorrectly classified as MAJ group as opposed to 11% of the MAJ group being incorrectly classified as LTC group.

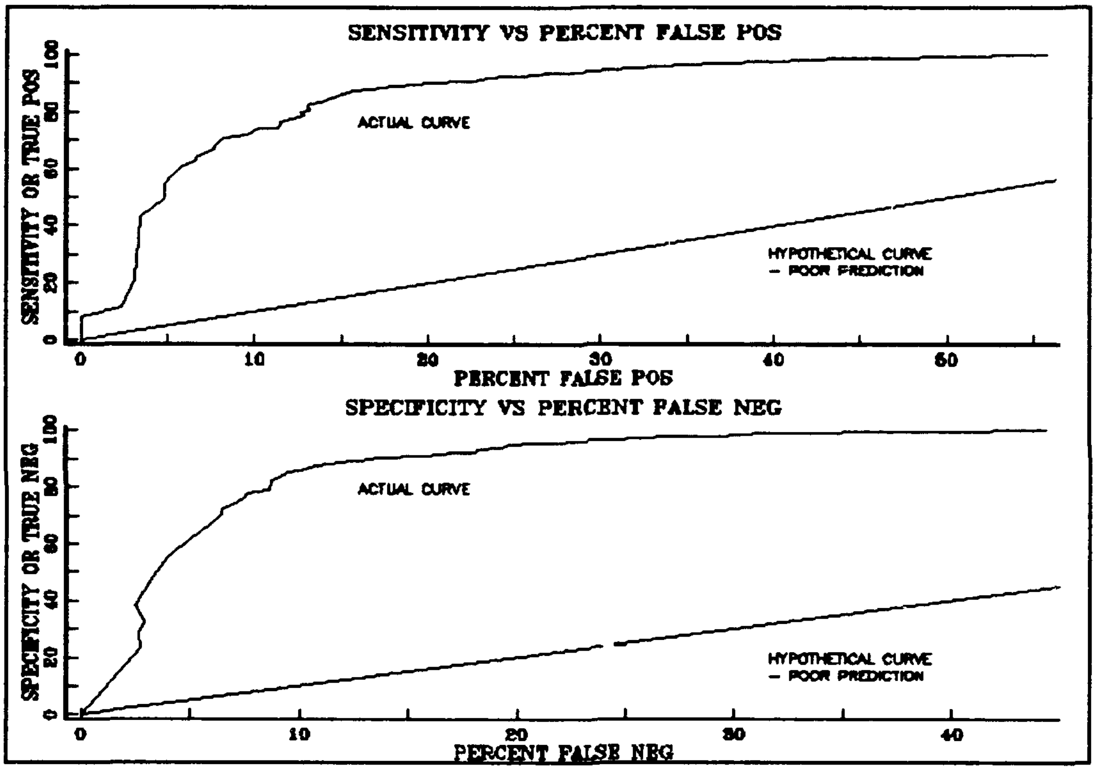

Figure 2: ROC Curves for CEP Binary Response Model.

2. Performance

For this model, Group I refers to officers who have a performance grade of B minus and below while Group II refers to those with a performance grade of at least a B. As can be seen from Figure 3, a cutoff point of about *0.64* would result in each

group being approximately *74%* correctly classified. On the contrary, a cutoff point of *0.2* would result in *98%* of Group I classified correctly but only about *28%* for Group II. Too high a cutoff point, for instance, a *0.8* cutoff value, would result in about *45%* of Group I classified correctly but about *91%* for Group II. Hence, proper choice of the cutoff value should be exercised so that each group would have a high percent of correct classification.

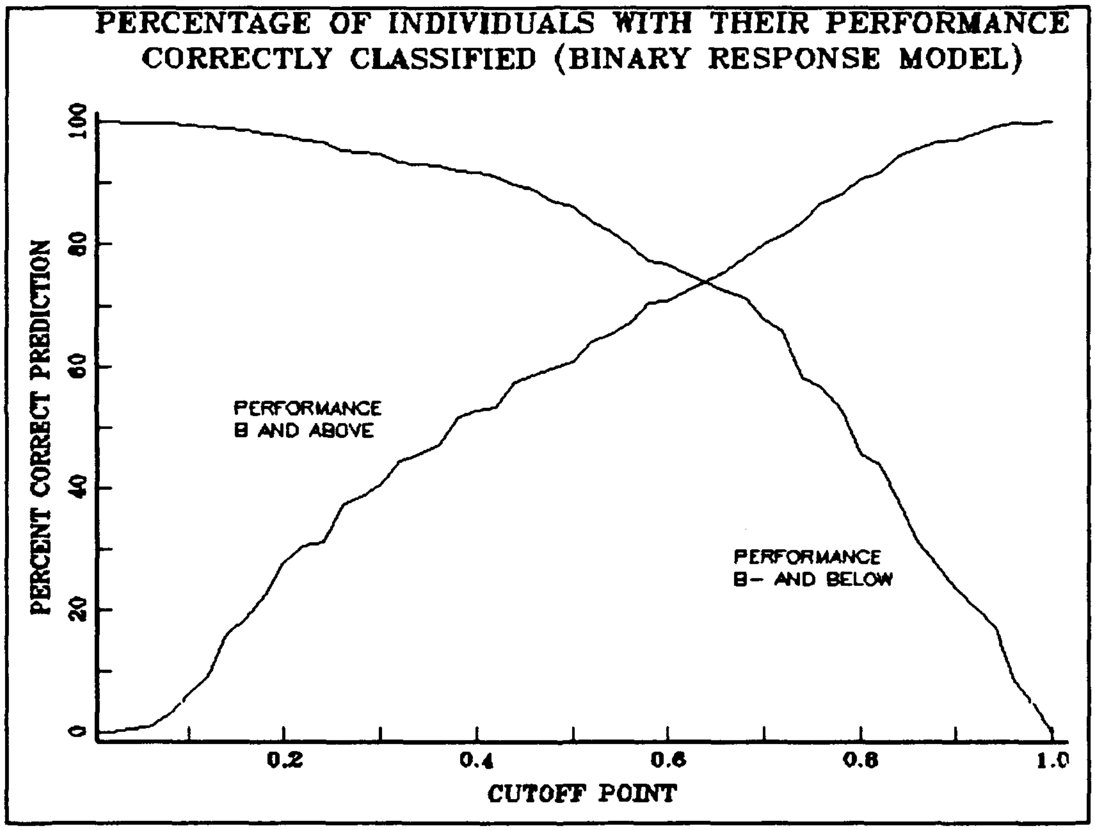

Figure 3: Percentage of Individuals with their Performance Correctly Classified (Binary Response Model).

From Figure 4, one can see clearly that the model derived does not give as good a prediction as the CEP model. A cutoff point of 0.64 would give about three

quarters of both groups being correctly classified with a corresponding false positive rate of *18%* and a false negative rate of *36%*. In other words, the percentage of individuals in Group I being incorrectly classified as Group II is twice that of Group II individuals being incorrectly classified as Group I.

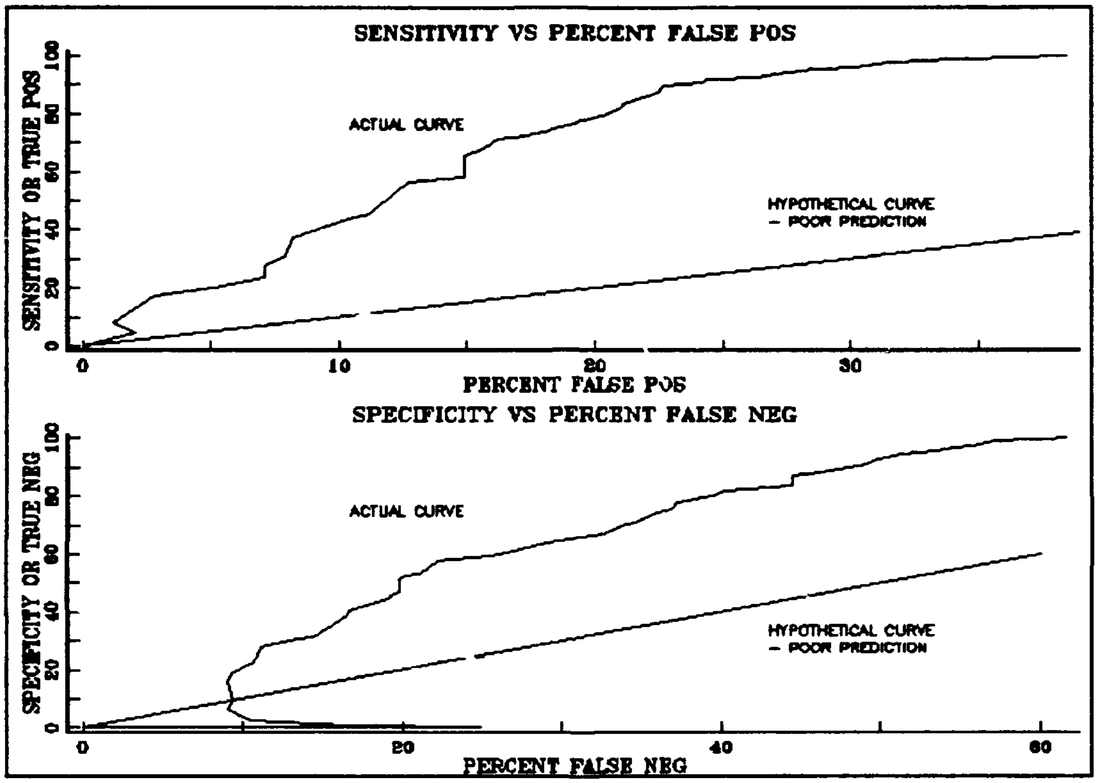

Figure 4: ROC Curves for Performance Binary Response Model.

The binary response model is the simplest model of the Linear Logistic Regression technique. In the following chapter, we will use a polytomous response model to consider response variables having more than two levels.

V. POLYTOMOUS RESPONSE MODEL

Valuable information are lost when the binary response models are used to model response variable having more than two levels. The numerous levels of the response variable (CEP and performance) are collapsed into two levels which are mutually exclusive. The power of the binary response model is realized when the response variable has two levels, as for example, officers being promoted or not promoted. Hence, for the CEP and performance models, it is essential to develop polytomous response models if more efficient discrimination of the officers is desired.

The candidate covariates considered in the model building are education level, training award, current rank, length of service, rank seniority, age, salary grade, previous year's (1991) annual performance grade and CEP estimate. The stepwise regression technique is again employed for variable selection. The significance levels for entry and staying in the model are set at α = 0.10 and 0.12 respectively. The cumulative logit model in SAS is used and it has the form

$$\log\left(\frac{\gamma_j(x)}{1-\gamma_j(x)}\right) = \theta_j + \beta^T x, \qquad j=1,\ldots,k-1 \quad , \tag{13}$$

where $\gamma_j(x) = pr(Y \leq j \mid x)$ is the cumulative probability up to and including category j, when the covariate vector is x. Refering to (4) on Page 11, the sign of $\beta^T x$ is opposite to that of (13) above. Hence, the signs of the parameter estimates obtained from the

SAS logistic procedure (using the cumulative logit model) must be reversed when (4) is used.

A. CURRENT ESTIMATED POTENTIAL

We will look first at Current Estimated Potential. The response variable is the CEP for the year 1992 and it has four levels - CPT, MAJ, LTC, and COL and above which are denoted by 1, 2, 3, and 4 respectively in the SAS program (see Appendix C, Section B). The resulting parameter estimates from SAS are given in Table 4 on the following page.

It is interesting to note that the set of covariates that entered the polytomous response model is the same as that for the binary response model. Further, the sign of the βs in the two models are the same.

The results show that as education level, current rank, rank seniority, previous year's annual performance grade and CEP estimate get higher, there is a tendency towards the higher-numbered categories. This means that it is more likely for the officer to have a high CEP estimate. Age, however, has the reverse effect.

B. PERFORMANCE

In the study of performance, the response variable is the annual performance grade for the year 1992. The original 15 levels (E^-,E,...,A,A^+) are collapsed to five levels representing A, B, C, D, and E grades (e.g., A^-,A, and A^+ are collapsed to form A, and so on). The SAS program can be found in Appendix C, Section B. The parameter estimates given by SAS are presented in Table 5.

Table 4. PARAMETER ESTIMATES FOR THE CEP MODEL

Variable	Parameter Estimate	Standard Error	Pr > Chi-Square
INTERCEP1	-0.2548	1.0019	0.7992
INTERCEP2	4.1478	1.0083	0.0001
INTERCEP3	9.7599	1.0848	0.0001
[1]EDU	-0.1952	0.0770	0.0112
[2]RANK	-2.2155	0.3302	0.0001
[3]RSNR	-0.1560	0.0557	0.0051
[4]AGE	0.2379	0.0439	0.0001
[5]P91	-0.1374	0.0560	0.0141
[6]C91	-1.2298	0.1176	0.0001

As in the case of the CEP study, the set of significant covariates that entered the polytomous response model is the same as that for the binary response model, but, of course the estimates are different for each model. Both the polytomous and binary response models give consistent results pertaining to the interpretation of the βs.

The results show that the more the number of years an officer remains in a particular rank and the higher the previous year's annual performance grade and CEP

[1]EDU is education level

[2]RANK is current rank

[3]RSNR is rank seniority

[4]Age is age of officer

[5]P91 is annual performance grade in the previous year (1991)

[6]C91 is CEP estimate in the previous year (1991)

estimate, the more likely it is for him to receive a high performance grade during the current year assessment. However, as an officer gets promoted to the next rank, there is a tendency for him to receive a poorer annual performance grade when compared to the grades he received before promotion. This could be a direct consequence for having quotas in the performance grades.

Table 5. PARAMETER ESTIMATE FOR THE PERFORMANCE MODEL

Variable	Parameter Estimate	Standard Error	Pr > Chi-Square
INTERCEP1	1.2353	0.4346	0.0045
INTERCEP2	2.0194	0.4213	0.0001
INTERCEP3	5.7756	0.4677	0.0001
INTERCEP4	9.8619	0.5844	0.0001
RANK	0.8447	0.1541	0.0001
RSNR	-0.2118	0.0333	0.0001
P91	-0.3637	0.0491	0.0001
C91	-0.5939	0.0879	0.0001

C. EVALUATION OF MODEL

It is useful to evaluate the models. To do this, the population is divided into two groups. The first group (Population I), is used for estimating the parameters while the second group (Population II) is used to assess the prediction quality of the model developed.

For an ordinal response, the LOGISTIC procedure in SAS performs a test of the parallel lines assumption. In the output, this test is labeled "Score Test for the Proportional Odds Assumption" when the logistic link function is selected. The null hypothesis is that the slope parameters are the same, against the alternative hypothesis that at least one pair of slope parameters are not the same.

1. Current Estimated Potential

The chi-square score from the statistical test for testing the proportional odds assumption, is 133.1061, which is significant with respect to a chi-square distribution with 12 degrees of freedom (*p=0.0001*). This indicates that a proportional odds model may not be so appropriate for the data. However, results show that the model developed has a *78 percent* correct prediction capability. When the model is tested on Population II, about *82 percent* of the officers in the group were classified correctly. Considering the fact that the model now has more information about the response variable (four levels as opposed to two levels for the binary response model), this is a reasonably good prediction model.

2. Performance

In the study of performance, the chi-square score for testing the proportional odds assumption, is 125.2833, which is again significant with respect to a chi-square distribution with 12 degrees of freedom (*p=0.0001*). The model is capable of correctly classifying about *68 percent* of the officers in both Population I and II. Not forgetting

that the response variable now has five levels, this model could be considered as being reasonably good.

In this and previous chapters, we have seen how the Logistic Regression technique may be used to estimate CEP and predict the performance grade of the officers. Next, we shall proceed to analyze the attrition behaviour of officers who entered service during the period from 1965-70 (denoted as the first cohort), 1971-76 (denoted as the second cohort), and 1977-82 (denoted as the third cohort).

VI. SURVIVAL ANALYSIS

This chapter compares and analyzes the attrition patterns of officers who entered service during the period 1965-70, 1971-76, and 1977-1982. Those officers who entered service before 1965 are not considered because there are only about a dozen of them. On the other hand, officers who entered service after 1982 are not considered because the number of years that can be studied, analyzed and compared are less than half of that in the first cohort (i.e., those who entered service during 1965-1970).

The attrition behaviour is analysed as a function of single covariate effect. The covariate effects considered are graduates against non-graduates, eduation (five levels), academic or overseas military training award against non-award holders, support vocations and service groups.

The Singapore military has a very young history. The military is formed after Singapore became independent in 1965. During the first few years, there are very few naval officers and pilots. Almost all the officers are in the Army. Hence, for the support vocations and sevice groups effects the study does not distinguished the various cohorts. Rather, a global view of the entire population is taken.

Graphical study of the survival functions is used for the comparative analyses. This approach gives a very good picture of how the various survival functions differ. The significance of the differences between survival functions are evaluated using formal statistical tests such as the Log-Rank and Wilcoxon test [Ref.5, Chap 5].

A. NON-GRADUATE AGAINST GRADUATE

The graduate group is defined as those officers who have attained at least an undergraduate degree. Survival functions for the three enlistment periods are shown in Figure 5. It is clear that there seems to be no significant difference between the non-graduate and graduate officers. The Log-Rank and Wilcoxon tests are both consistent with this visual observation.

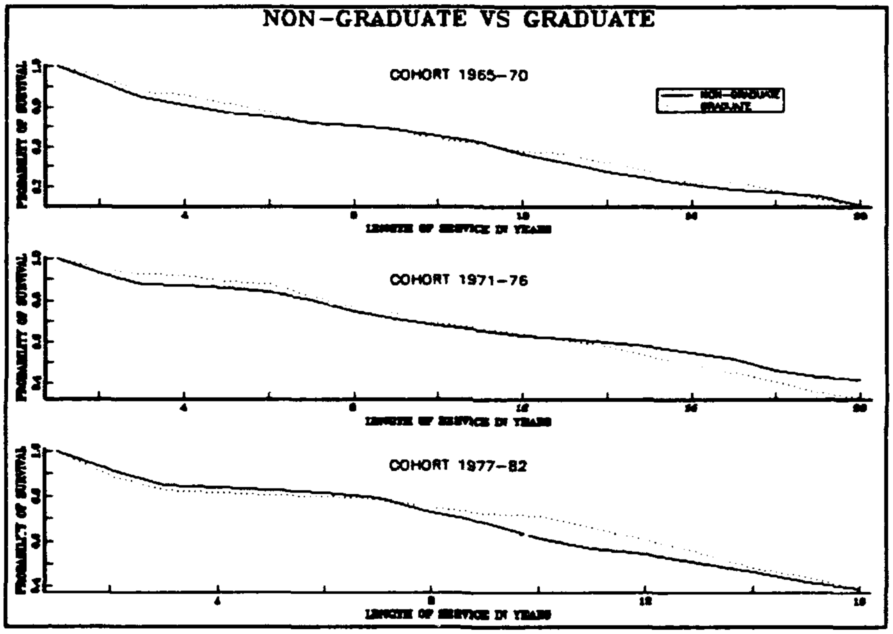

Figure 5: Survival Curves for Non-Graduates and Graduates.

B. EDUCATIONAL LEVEL

Figure 6 shows survival curves for various education levels. The 'O-' and 'A-' levels represent officers who have a GCE 'O-' and 'A-' level respectively. 'Diploma' represents officers who have only an Advanced or Basic Diploma education.

'Undergrad' denotes officers who have an Undergraduate Degree. 'Postgrad' denotes officers who have a Postgraduate Degree.

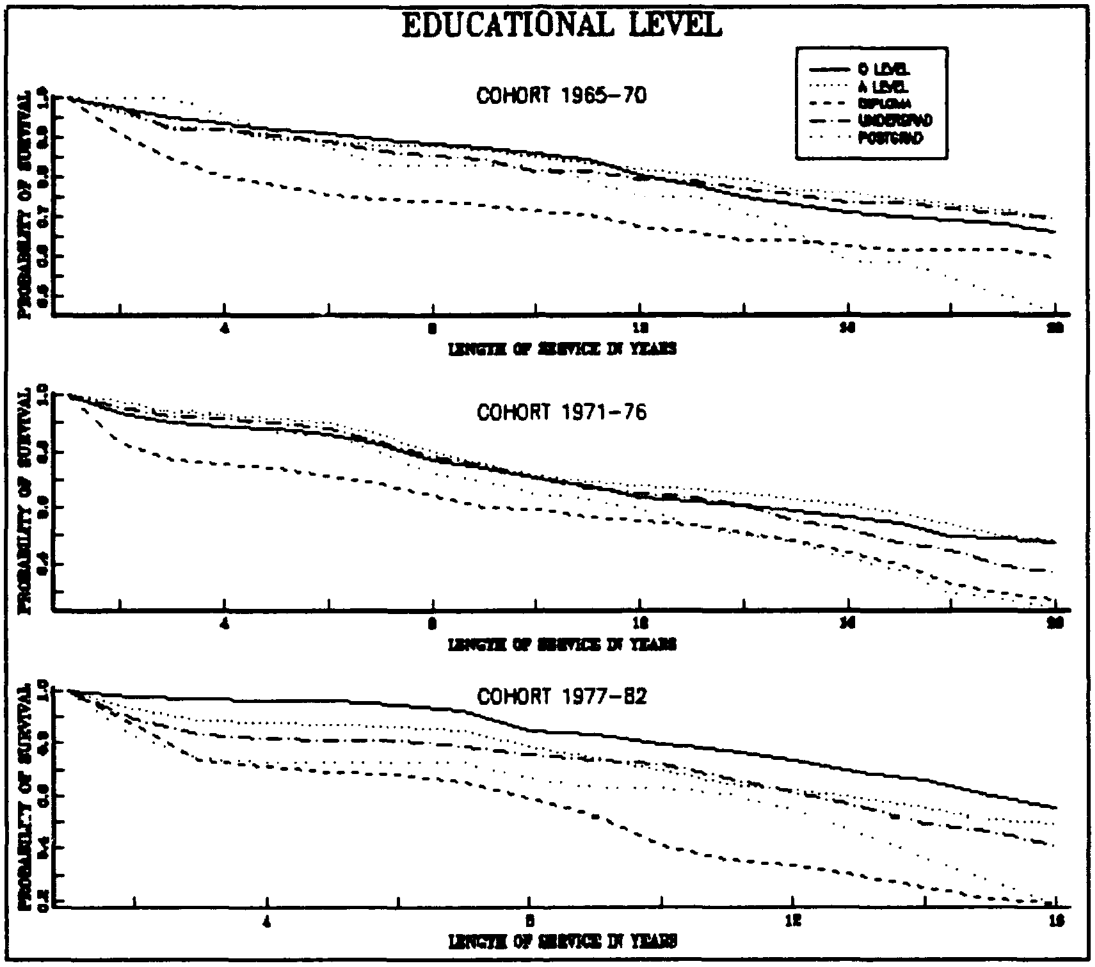

Figure 6: Survival Curves for Different Education Levels.

It is interesting to observe from Figure 6 that officers with an 'O-' or 'A-' level education have consistently survived longer in service than the others for all the three cohorts. On the contrary, officers with diploma education show consistently the lowest survival function. For this group of officers, it can be seen that there is a sharp drop

in the survival function for the first two to three years of service. After which it decreases more or less in a steady manner. The only exception is for Cohort 3 (1977-82) where there is again another sharp drop in the survival function after about nine to ten years in service. In terms of survival function, the officers with undergraduate degrees seem to rank below the 'O-' and 'A-' level officers but above those with postgraduate degrees.

From the top plot of Figure 6 it appears that except for the officers with diploma and postgraduate education, the survival functions of the remaining groups of officers seem to be more a less the same. This suspicion is confirmed by examining the Log-Rank (*p-value = 0.031*) and Wilcoxon (*p-value = 0.0108*) tests for Cohort 1 (1965-70). Both of these tests give *p-value of 0.0001* for the other two cohorts indicating a strong significant difference in attrition behaviour among different education levels. The Log-Rank and Wilcoxon tests are recomputed without the officers with diploma qualification. It is found that for Cohort 1, education level is not a significant covariate at the 0.05 significance level. Here, the Log-Rank test *p-value is 0.0805*, and the Wilcoxon test *p-value is 0.1109*. For cohorts 2 (1971-76) and 3 (1977-82), however, education level is again found to be a significant covariate.

From the foregoing disscusions it can be concluded that there is a significant difference in attrition behaviour between officers with diploma education and those with other educational qualifications. As for the other education levels ('0-' and 'A-' levels, 'under-' and 'post-'graduates) the survival function seems to indicate towards a strong significant difference among differing education levels. However, a note of caution is

that the attrition behaviour is a function of many other complex and uncontrollable factors such as civilian job market opportunities, the country's economy, inflation, unemployment rates, etc. In other words, the trend of the survival functions should be viewed with caution.

C. NON-AWARD AGAINST TRAINING AWARD HOLDERS

Officers who are given academic or overseas military training awards are expected to survive longer in service than those who are not. One simple reason being officers given awards are required to sign an obligated service contract of between five to eight years, depending on the type of training award they received. If the officer breaks this contract, he would have to reimburse the Government the money invested in him. The survival functions are shown in Figure 7.

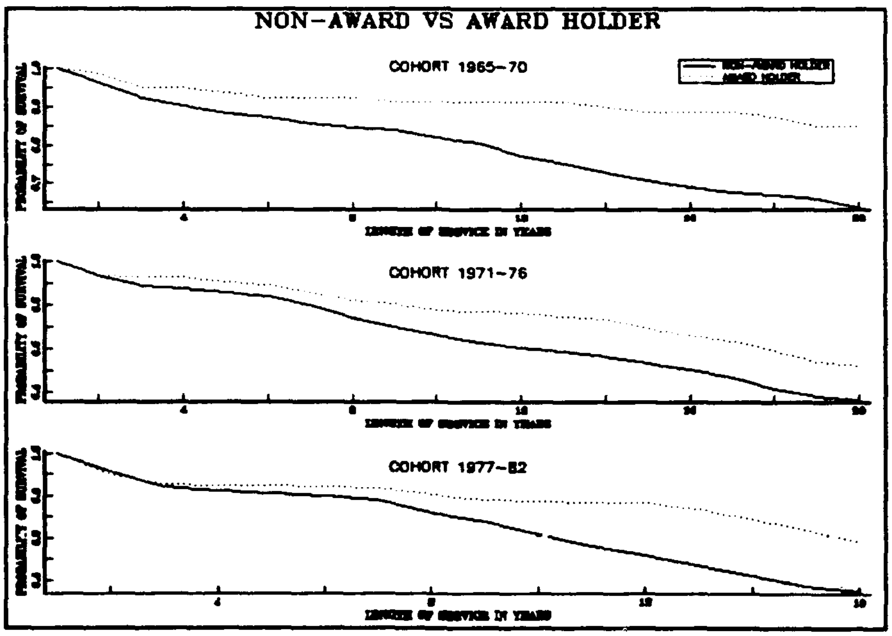

Figure 7: Survival Curves for Non-Award and Award Holders.

The difference in the survival functions between the two groups of officers are roughly the same for the three cohorts. This indicates a very strong consistency in the the attrition behaviour for the three cohorts. Figure 8 shows the plot of the difference in survival functions between this two groups of officers for the three cohorts.

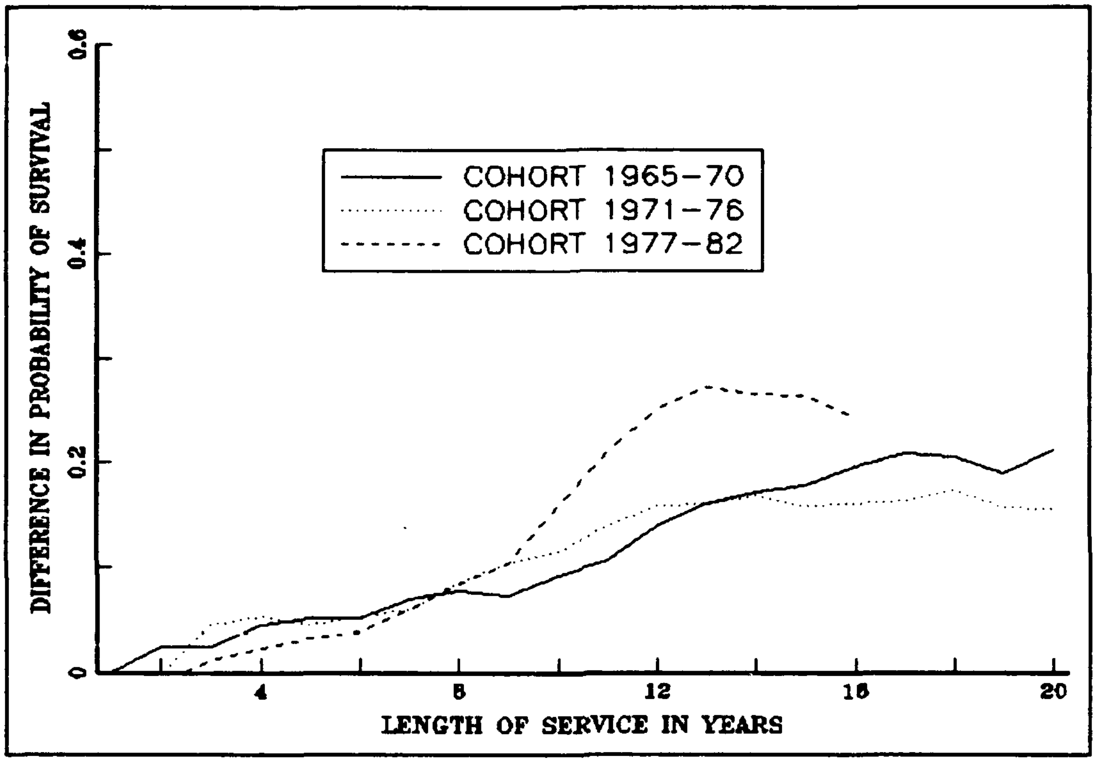

Figure 8: Difference in Survival Funtions Between Non-Award and Award Holders.

D. SUPPORT VOCATION

This study includes Engineering officers, Army and Air Force support officers. The Engineering category consists of Ordnance, Electric, Naval and Air Engineering officers. The Army support consists of Signal, Artillery, Mechanical Transport, Armour

'Reccee' and Armour Infantry officers. The Air Force support consists of Air Defence and Air Operations & Communication officers.

As shown in Figure 9, the survival function of the Army support officers exhibits an almost linear trend which suggests a constant attrition rate. The survival functions of the Engineering and Air Force support officers could be pooled and described by a single two piece-wise linear functions since their attrition behaviours are roughly the same. For the first three years in service, both these two groups of officers show a very sharp drop in the survival function compared with that of the Army support officers. After the third year of service, the slopes of the survival functions for the three categories of officers are more a less the same.

Figure 10 shows the hazard function estimates of the above three categories of officers. The attrition rate is the highest in the first year of service for the Engineering and Air Force support officers, and drops to the lowest at the beginning of the third year. After the third year the attrition rate of the Engineering officers is generally higher than the other two categories of officers. On the contrary, the Army support officers exhibit a relatively constant attrition rate throughout the entire period of study.

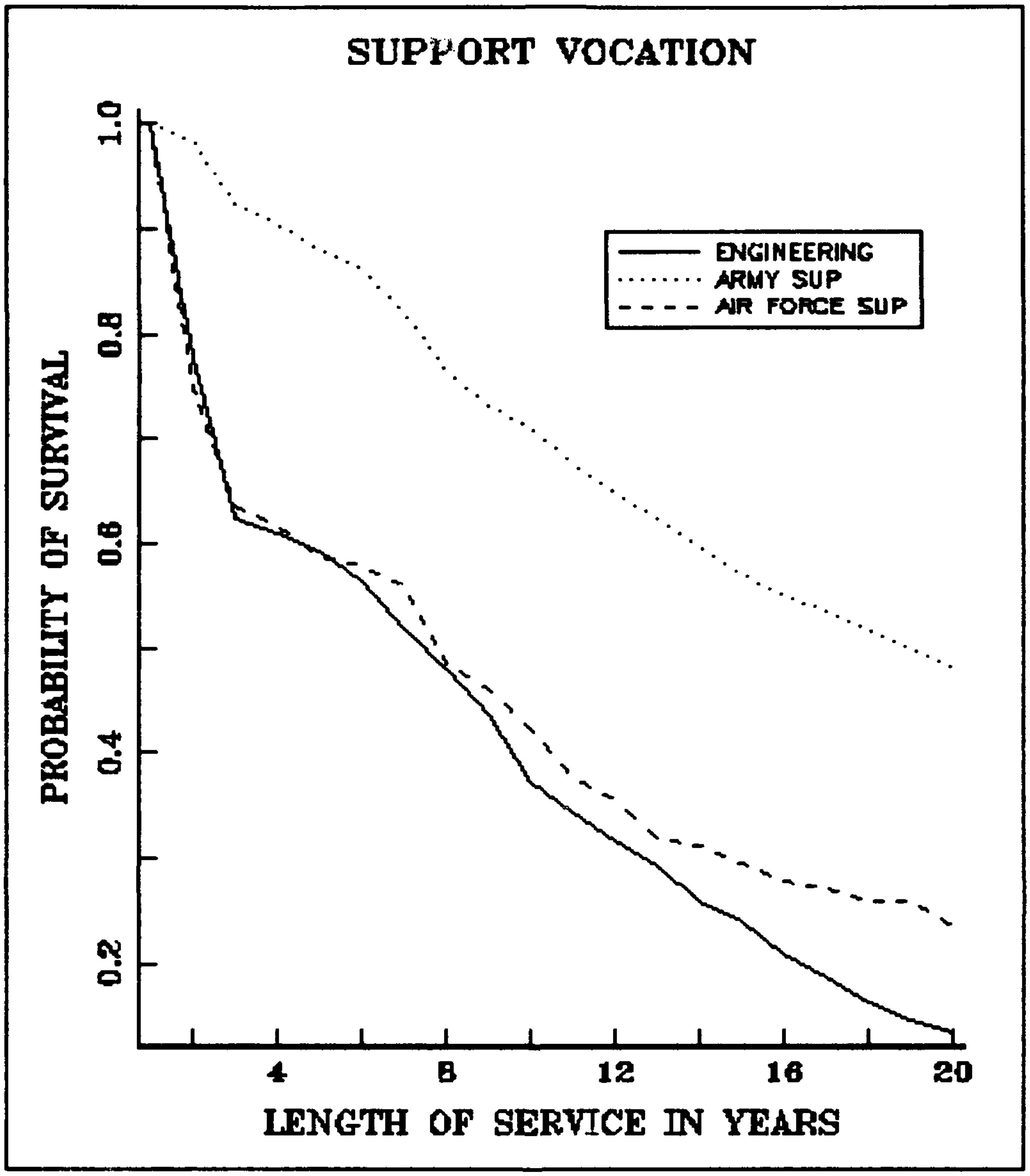

Figure 9: Survival Curves for Three Support Vocations.

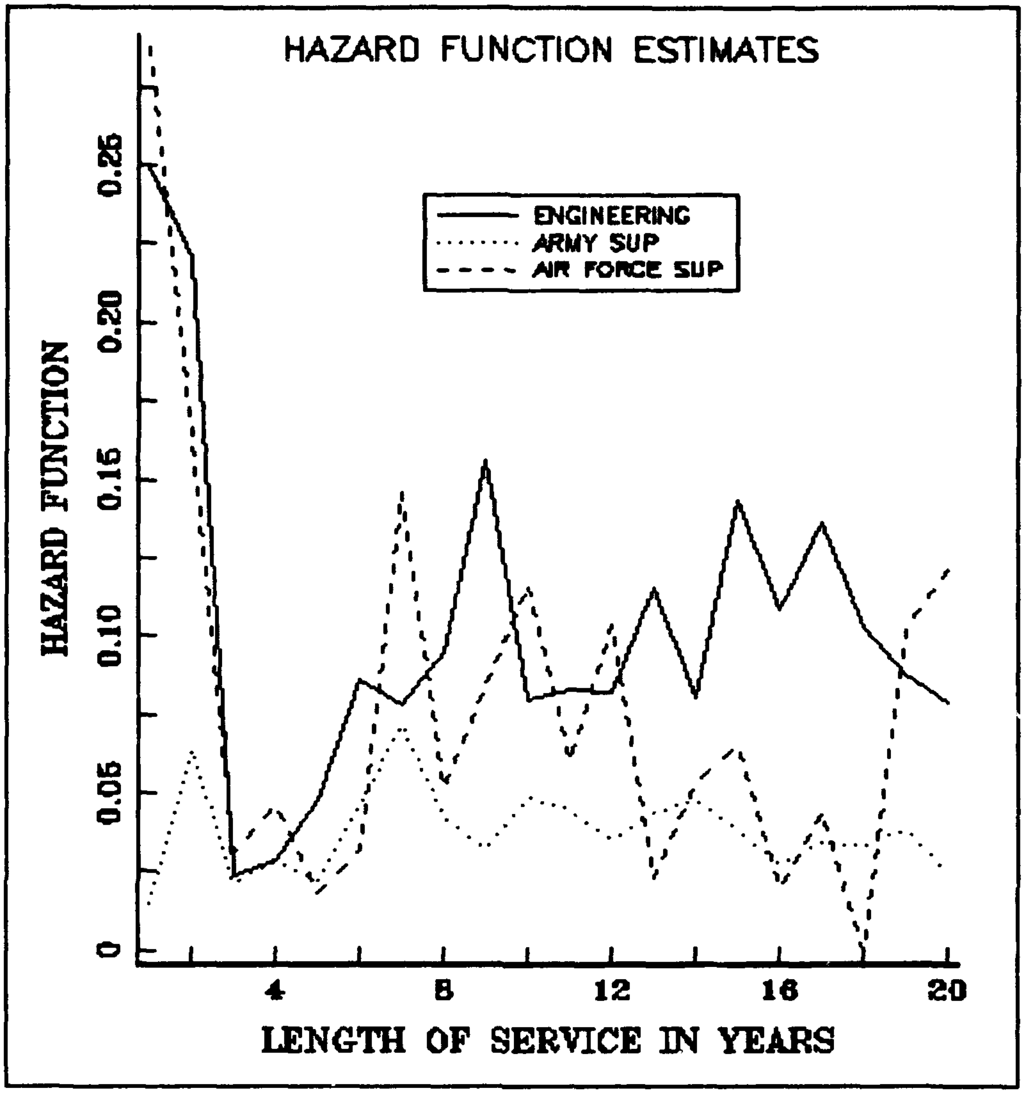

Figure 10: Hazard Function Estimates for Three Support Vocations.

E. SERVICE GROUPS

The three groups of service under study are Infantry and Guards (Army), Pilots (Air Force), and Naval (Navy) officers. The pilots are either on the pensionable or 12 years contract scheme. Therefore, it is not surprising to find that they have the best

survival among the service groups (see Figure 11) and that their attrition rate begins to escalate only after 12 years of service (see Figure 12).

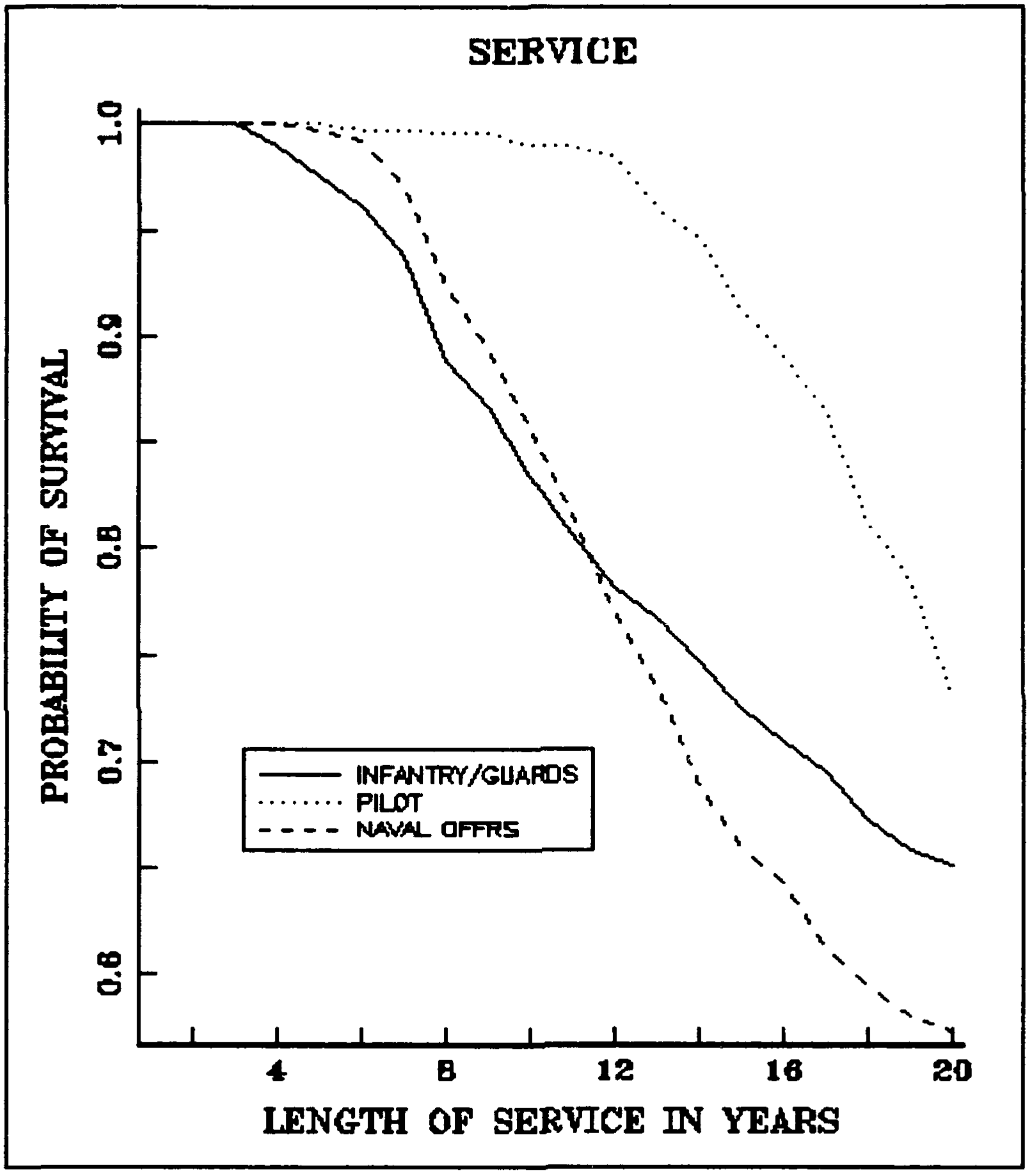

Figure 11: Survival Curves for Different Service Groups.

The highest attrition rate occurs at year six for both the Army and Naval officers because of their six years contract, as opposed to the pilots who have 12 years contract. For the first six years of service, the Naval officers have a lower risk of leaving the service than their Army counterparts. After the first six years of service, the converse is true.

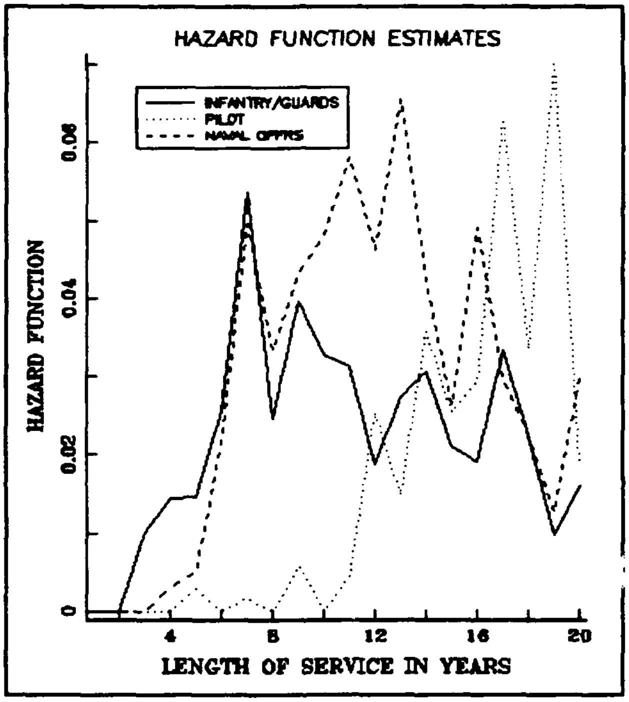

Figure 12: Hazard Function Estimates for Service Groups.

In this chapter, we have seen how the Survival Analysis technique may be used to analyze the attrition behaviours of the officers in the Singapore military. The following chapter gives the conclusions and summary of these and earlier findings, together with recommendations for future work.

VII. CONCLUSIONS

A. LOGISTIC REGRESSION ANALYSIS

The Logistic Regression technique is frequently used for analysis of data collected retrospectively. It is commonly used when an individual is to be classified into two or more categories. The amalgamation of response categories to two levels results in the lost of valuable information, and is discouraged if efficient discrimination of the response categories is desired.

The significant results of the study on CEP estimation and performance prediction are briefly outlined below.

- **Education Level-** Education level is not a significant predictor of performance though a higher education level seems to give an indication of higher CEP.

- **Training Award-** There is insufficient evidence to support the notion that officers given an academic or overseas military training award tends to have a better performance grade than those who did not receive any.

- **Rank-** The higher the rank of an officer, the more likely it is for him to get a poorer performance grade than when he was in the previous rank.

- **Previous year's CEP and Performance Grade-** Current year's CEP estimation and performance grade prediction are highly correlated to previous year's CEP and performance grade.

B. SURVIVAL ANALYSIS

An intrinsic characteristic of survival data is the presence of censored observations. It would be impractical to wait until every subject has "died" before conducting any analysis. The life-table or product-limit estimate of the survival function is an invaluable tool to analyze the attrition behaviour when censored observations are present in the data set.

The graphical approach of analyzing the survival function is a simple way of analyzing the problem without the requirement of a statistics background. Although some of the results are trivial, the analysis gives a clear insight on the attrition behaviour of the officers who entered service during the three enlistment periods (1965-70, 1971-76, and 1977-82). The results of the analysis are briefly outlined below.

- **Non-Graduate vs Graduate-** For each of the three enlistment periods the attrition behaviour between non-graduates and graduates is not significantly different.

- **Education Level-** Education level has a strong relationship with the attrition behaviour of the officers. Officers with an 'O-' or 'A-' level qualification have consistently survived longer in the service than officers who have any other educational qualifications. On the contrary, officers with diploma qualification exhibit the lowest survival functions.

- **Training Award-** The trend of the difference in the survival functions between non-award and award holders for the three enlistment groups is statistically the same.

- **Support Vocation-** The Engineering and Air Force support officers have the highest attrition rate during the first year of service. It drops to the lowest at the beginning of the third year, after which the attrition rates of the Engineering officers are generally higher than the other two categories of officers. The Army support officers exhibit a relatively constant attrition rate throughout the entire period of study.

- **Service Group-** For the first six years of service, the Naval officers have a lower risk of leaving the service than their Army counterparts. In contrast, after the first six years period, the converse is true.

C. RECOMMENDATIONS FOR FUTURE STUDY

Data on the officer's extra-curriculum activities during his school days, marital status, number of children, and the Officer Cadet School's graduation grade are some of the interesting covariates that could be investigated in future studies.

Having analyzed the attrition behaviour of the officers the next step would be to predict the number of officers in each rank leaving the service based on Singapore's economic indicators (e.g., unemployment rate, inflation, gross national product, etc.).

Another interesting area to look at is to check whether there is any significant difference in performance and CEP among officers of different vocations.

It is hoped that the models developed in this thesis and the insights they provide will be beneficial to manpower planners and recruitment agencies.

LIST OF REFERENCES

1. Agresti, A., Categorical Data Analysis, John Wiley & Sons, Inc., 1990.

2. David W. Hosmer, and Stanley Lemeshow, Applied Logistic Regression, John Wiley & Sons, Inc., 1989.

3. Afifi, A.A., and Virginia Clark, Computer-Aided Multivariate Analysis, 2d ed., Van Nostrand Reinhold Company, 1990.

4. McCullagh P., and Nelder FRS, J.A., Generalized Linear Models, 2d ed., Chapman & Hall, 1989.

5. Elisa, T. Lee, Statistical Methods for Survival Data Analysis, 2d ed., John Wiley & Sons, Inc., 1992.

6. Kaplan, E.L., and Meier, P., "Nonparametric Estimation from Incomplete Observations," Journal of the American Statistical Association, v.53, pp. 457-481, 1958.

7. SAS/STAT User's Guide, Volume 2, GLM-VARCOMP, Version 6, Fourth Edition.

APPENDIX A: ADDITIONAL TWO-WAY TABLES OF CEP AND PERFORMANCE

TABLE 6: TABLE OF EDUCATION LEVEL BY CEP FOR THE YEAR 1992

EDUCATION	CEP 1992					
PERCENT ROW PCT COL PCT	CPT	MAJ	LTC	COL	BG, MG	TOTAL
NON-GRADUATE	29.68 40.49 92.12	25.47 34.75 88.54	17.70 24.15 51.49	0.45 0.62 10.08	0.00 0.00 0.00	73.31
GRADUATE	2.54 9.52 7.88	3.30 12.36 11.46	16.68 62.50 48.51	4.06 15.20 89.92	0.11 0.43 100.00	26.69
TOTAL	32.22	28.77	34.38	4.51	0.11	100.00

Statistic	DF	Value	Prob
Chi-Square	4	715.763	0.000
Likelihood Ratio Chi-Square	4	716.807	0.000
Mantel-Haenszel Chi-Square	1	608.738	0.000
Phi Coefficient		0.521	
Contingency Coefficient		0.462	
Cramer's V		0.521	

TABLE 7: TABLE OF TRAINING AWARD BY CEP FOR THE YEAR 1992

AWARD	CEP 1992					
PERCENT ROW PCT COL PCT	CPT	MAJ	LTC	COL	BG, MG	TOTAL
NON-AWARD HOLDER	18.20 28.10 56.47	26.12 40.34 90.78	19.86 30.68 57.77	0.57 0.88 12.61	0.00 0.00 0.00	64.75
TRAINING AWARD HOLDER	14.03 39.78 43.53	2.65 7.53 9.22	14.52 41.18 42.23	3.94 11.18 87.39	0.11 0.32 100.00	35.25
TOTAL	32.22	28.77	34.38	4.51	0.11	2638 100.00

Statistic	DF	Value	Prob
Chi-Square	4	417.398	0.000
Likelihood Ratio Chi-Square	4	467.530	0.000
Mantel-Haenszel Chi-Square	1	29.695	0.000
Phi Coefficient		0.398	
Contingency Coefficient		0.370	
Cramer's V		0.398	

TABLE 8: TABLE OF LENGTH OF SERVICE BY CEP FOR THE YEAR 1992

LENGTH OF SERVICE	CEP 1992					
PERCENT ROW PCT COL PCT	CPT	MAJ	LTC	COL	BG, MG	TOTAL
≤ 6	27.45 66.36 85.18	11.14 26.95 38.74	2.73 6.60 7.94	0.04 0.09 0.84	0.00 0.00 0.00	41.36
7 TO ≤ 12	3.18 9.31 9.88	10.61 31.04 36.89	18.23 53.33 53.03	2.16 6.32 47.90	0.00 0.00 0.00	34.19
≥ 13	1.59 6.51 4.94	7.01 28.68 24.37	13.42 54.88 39.03	2.31 9.46 51.26	0.11 0.47 100.00	645 24.45
TOTAL	32.22	28.77	34.38	4.51	0.11	100.00

Statistic	DF	Value	Prob
Chi-Square	8	1192.703	0.000
Likelihood Ratio Chi-Square	8	1351.217	0.000
Mantel-Haenszel Chi-Square	1	934.707	0.000
Phi Coefficient		0.672	
Contingency Coefficient		0.558	
Cramer's V		0.475	

TABLE 9: TABLE OF RANK SENIORITY BY CEP FOR THE YEAR 1992

RANK SENIORITY	CEP 1992					
PERCENT ROW PCT COL PCT	CPT	MAJ	LTC	COL	BG, MG	TOTAL
≤ 3	27.90 40.69 86.59	18.23 26.59 63.37	19.07 27.81 55.46	3.30 4.81 73.11	0.08 0.11 66.67	68.57
4 TO ≤ 6	2.84 11.96 8.82	7.73 32.54 26.88	11.98 50.40 34.84	1.18 4.94 26.05	0.04 0.16 33.33	23.77
≥ 7	1.48 19.31 4.59	2.81 36.63 9.75	3.34 43.56 9.70	0.04 0.50 0.84	0.00 0.00 0.00	7.66
TOTAL	32.22	28.77	34.38	4.51	0.11	100.00

Statistic	DF	Value	Prob
Chi-Square	8	223.649	0.000
Likelihood Ratio Chi-Square	8	247.516	0.000
Mantel-Haenszel Chi-Square	1	96.038	0.000
Phi Coefficient		0.291	
Contingency Coefficient		0.280	
Cramer's V		0.206	

TABLE 10: TABLE OF AGE GROUP BY CEP FOR THE YEAR 1992

AGE GROUP	CEP 1992					
PERCENT ROW PCT COL PCT	CPT	MAJ	LTC	COL	BG, MG	TOTAL
≤ 25	26.38 71.31 81.88	7.77 21.00 27.01	2.65 7.17 7.72	0.19 0.51 4.20	0.00 0.00 0.00	37.00
26 TO ≤ 30	4.06 11.53 12.59	12.55 35.67 43.61	16.60 47.20 48.29	1.97 5.60 43.70	0.00 0.00 0.00	35.18
31 TO ≤ 35	1.02 4.66 3.18	6.48 29.48 22.53	12.55 57.07 36.49	1.82 8.28 40.34	0.11 0.52 100.00	21.99
36 TO ≤ 40	0.19 5.68 0.59	1.18 35.23 4.08	1.59 47.73 4.63	0.38 11.36 8.40	0.00 0.00 0.00	3.34
41 TO ≤ 45	0.42 18.97 1.29	0.72 32.76 2.50	0.95 43.10 2.76	0.11 5.17 2.52	0.00 0.00 0.00	2.20
≥ 46	0.15 50.00 0.47	0.08 25.00 0.26	0.04 12.50 0.11	0.04 12.50 0.84	0.00 0.00 0.00	0.30
TOTAL	32.22	28.77	34.38	4.51	0.11	100.00

Statistic	DF	Value	Prob
Chi-Square	20	1208.219	0.000
Likelihood Ratio Chi-Square	20	1314.311	0.000
Mantel-Haenszel Chi-Square	1	663.447	0.000
Phi Coefficient		0.677	
Contingency Coefficient		0.560	
Cramer's V		0.338	

TABLE 11: TABLE OF EDUCATION LEVEL BY PERFORMANCE FOR THE YEAR 1992

EDUCATION	PERFORMANCE 1992					
PERCENT ROW PCT COL PCT	E	D	C	B	A	TOTAL
NON- GRADUATE	27.75 37.85 88.51	2.69 3.67 80.68	27.18 37.07 67.20	14.94 20.37 63.45	0.76 1.03 57.14	73.31
GRADUATE	3.60 13.49 11.49	0.64 2.41 19.32	13.27 49.72 32.80	8.61 32.24 36.55	0.57 2.13 42.86	26.69
TOTAL	31.35	3.34	40.45	23.54	1.33	100.00

Statistic	DF	Value	Prob
Chi-Square	4	156.071	0.000
Likelihood Ratio Chi-Square	4	170.985	0.000
Mantel-Haenszel Chi-Square	1	149.270	0.000
Phi Coefficient		0.243	
Contingency Coefficient		0.236	
Cramer's V		0.243	

TABLE 12: TABLE OF TRAINING AWARD BY PERFORMANCE FOR THE YEAR 1992

AWARD	PERFORMANCE 1992					
PERCENT ROW PCT COL PCT	E	D	C	B	A	TOTAL
NON-AWARD HOLDER	16.11 24.88 51.39	2.96 4.57 88.64	28.96 44.73 71.60	15.85 24.47 67.31	0.87 1.35 65.71	64.75
TRAINING AWARD HOLDER	15.24 43.23 48.61	0.38 1.08 11.36	11.49 32.58 28.40	7.70 21.83 32.69	0.45 1.29 34.29	35.25
TOTAL	31.35	3.34	40.45	23.54	1.33	100.00

Statistic	DF	Value	Prob
Chi-Square	4	110.410	0.000
Likelihood Ratio Chi-Square	4	112.829	0.000
Mantel-Haenszel Chi-Square	1	54.901	0.000
Phi Coefficient		0.205	
Contingency Coefficient		0.200	
Cramer's V		0.205	

TABLE 13: TABLE OF LENGTH OF SERVICE BY PERFORMANCE FOR THE YEAR 1992

LENGTH OF SERVICE	PERFORMANCE 1992					
PERCENT ROW PCT COL PCT	E	D	C	B	A	TOTAL
≤ 6	26.76 64.71 85.37	0.87 2.11 26.14	10.84 26.21 26.80	2.88 6.97 12.24	0.00 0.00 0.00	41.36
7 TO ≤ 12	3.18 9.31 10.16	1.14 3.33 34.09	18.04 52.77 44.61	11.22 32.82 47.67	0.61 1.77 45.71	34.19
≥ 13	1.40 5.74 4.47	1.33 5.43 39.77	11.56 47.29 28.58	9.44 38.60 40.10	0.72 2.95 54.29	24.45
TOTAL	31.35	3.34	40.45	23.54	1.33	100.00

Statistic	DF	Value	Prob
Chi-Square	8	1022.428	0.000
Likelihood Ratio Chi-Square	8	1104.442	0.000
Mantel-Haenszel Chi-Square	1	784.896	0.000
Phi Coefficient		0.623	
Contingency Coefficient		0.529	
Cramer's V		0.440	

TABLE 14: TABLE OF RANK SENIORITY BY PERFORMANCE FOR THE YEAR 1992

RANK SENIORITY	PERFORMANCE 1992					
PERCENT ROW PCT COL PCT	E	D	C	B	A	TOTAL
≤ 3	28.24 41.18 90.08	2.12 3.10 63.64	28.73 41.90 71.04	9.17 13.38 38.97	0.30 0.44 22.86	68.57
4 TO ≤ 6	2.65 11.16 8.46	0.49 2.07 14.77	9.29 39.07 22.96	10.58 44.50 44.93	0.76 3.19 57.14	23.77
≥ 7	0.45 5.94 1.45	0.72 9.41 21.59	2.43 31.68 6.00	3.79 49.50 16.10	0.27 3.47 20.00	7.66
TOTAL	31.35	3.34	40.45	23.54	1.33	100.00

Statistic	DF	Value	Prob
Chi-Square	8	497.816	0.000
Likelihood Ratio Chi-Square	8	507.451	0.000
Mantel-Haenszel Chi-Square	1	353.103	0.000
Phi Coefficient		0.434	
Contingency Coefficient		0.398	
Cramer's V		0.307	

TABLE 15: TABLE OF AGE GROUP BY PERFORMANCE FOR THE YEAR 1992

AGE GROUP	PERFORMANCE 1992					
PERCENT ROW PCT COL PCT	E	D	C	B	A	TOTAL
≤ 25	26.16 70.70 83.43	0.72 1.95 21.59	8.68 23.46 21.46	1.36 3.69 5.80	0.08 0.20 5.71	37.00
26 TO ≤ 30	3.68 10.45 11.73	1.06 3.02 31.82	18.69 53.13 46.20	11.30 32.11 47.99	0.45 1.29 34.29	35.18
31 TO ≤ 35	1.18 5.34 3.75	0.95 4.31 28.41	10.05 45.69 24.84	9.14 41.55 38.81	0.68 3.10 51.43	21.99
36 TO ≤ 40	0.30 9.09 0.97	0.19 5.68 5.68	1.78 53.41 4.40	0.99 29.55 4.19	0.08 2.27 5.71	3.34
41 TO ≤ 45	0.04 1.72 0.12	0.34 15.52 10.23	1.02 46.55 2.53	0.76 34.48 3.22	0.04 1.72 2.86	2.20
≥ 46	0.00 0.00 0.00	0.08 25.00 2.27	0.23 75.00 0.56	0.00 0.00 0.00	0.00 0.00 0.00	0.30
TOTAL	31.35	3.34	40.45	23.54	1.33	100.00

Statistic	DF	Value	Prob
Chi-Square	20	1234.393	0.000
Likelihood Ratio Chi-Square	20	1304.519	0.000
Mantel-Haenszel Chi-Square	1	694.047	0.000
Phi Coefficient		0.684	
Contingency Coefficient		0.565	
Cramer's V		0.342	

APPENDIX B: COMPARISON OF BINARY RESPONSE MODELS

A sensitivity analysis is carried out to determine the various outcomes derived from the three sets of covariate combinations using the same data set. The variables that entered the model should be reasonable and practical besides being the best fitting covariates.

Before proceeding further, it is necessary to discuss the various statistics that are used to assess the model fit. The Akaike Information Criterion (AIC) and Schwartz Criterion (SC) statistics under "Criteria for Assessing Model Fit" (see the example of SAS output in Appendix C) are primarily used for comparing different models for the same data. In general, when comparing models, lower values of these two statistics indicate a better model. [Ref. 7:p. 1088]

The Score statistic gives a test for the joint significance of the explanatory variables in the model. This test considers only the independent variables, so no test is shown for the columns for "Intercept Only" and "Intercept and Covariates." The -2 LOG L row gives statistics and a test for the effects of the covariates based on -2 Log Likelihood (see Pages 65, 68, 71, 75, 78 and 81).

A. CURRENT ESTIMATED POTENTIAL

The SAS outputs (Appendix C) for the three models indicate that the most desirable model for CEP estimation is Model 1 (AIC: 868.336; SC: 904.168), followed by Model 2 (AIC: 1124.762; SC: 1162.575) and model 3 (AIC: 2028.528; SC: 2057.915). However, a closer look at the parameter estimates of Model 1 shows evidence that multicollinearity may exist. The parameter estimates for performance grade for the previous one and two years are of different signs (P91: -0.1831, P90: 0.1456) indicating opposite effect for the same unit change in performance grade. This does not seem to make sense. Since the performance grades in the previous two years are likely to be highly intercorrelated, the computed estimates of the regression coefficients are unstable and their interpretation becomes tenuous. Hence, Model 2 is selected.

The reader would appreciate much better by referring to Figure 13 on the following page. The top graph gives the plot of senstivity against percent false positive rate for the three models under consideration. As seen, the three models are marginally different from each other since the three curves in the top plot are relatively close to each other. Although Model 3 outperforms marginally (for Percent False POS > 7) than the other two models for CEP prediction of the MAJ* group, it has much poorer prediction power for CEP of the LTC** group (see bottom graph of Figure 13).

B. PERFORMANCE

Once again, Model 1 proves to be the most statistically desirable model if one compares the AIC and SC statistics of the three models. However, why should performance depend on C91 and C89, but not C90? All these three variables measure the same characteristic (i.e., CEP but in three consecutive years). Although CEP estimation is supposed to be conducted independently from year to year, we cannot discount totally the fact that there may be some intercorrelation. Hence, Model 2 is selected instead.

The top and bottom graphs in Figure 14 show the plots of sensitivity against percent false positive rate, and specificity against percent false negative rate respectively. Again, Model 3 outperforms the other two models for performance prediction of Group I+, but it is almost useless for prediction of Group II++, as seen by the large portion of the graph falling below the hypothetical curve.

---

* Population with CEP of Senior MAJ and below

** Population with CEP of LTC and above

\+ Population with performance grade of B minus and below

++ Population with performance grade of B and above

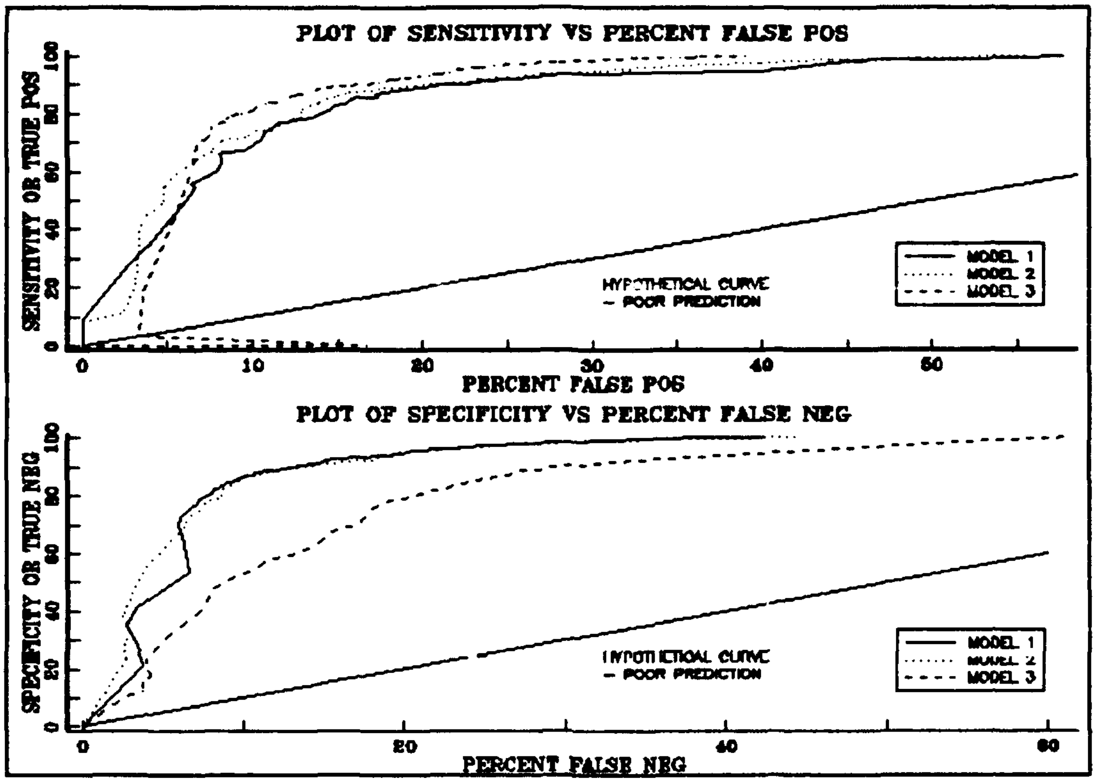

Figure 13: Comparison of CEP Binary Response Models.

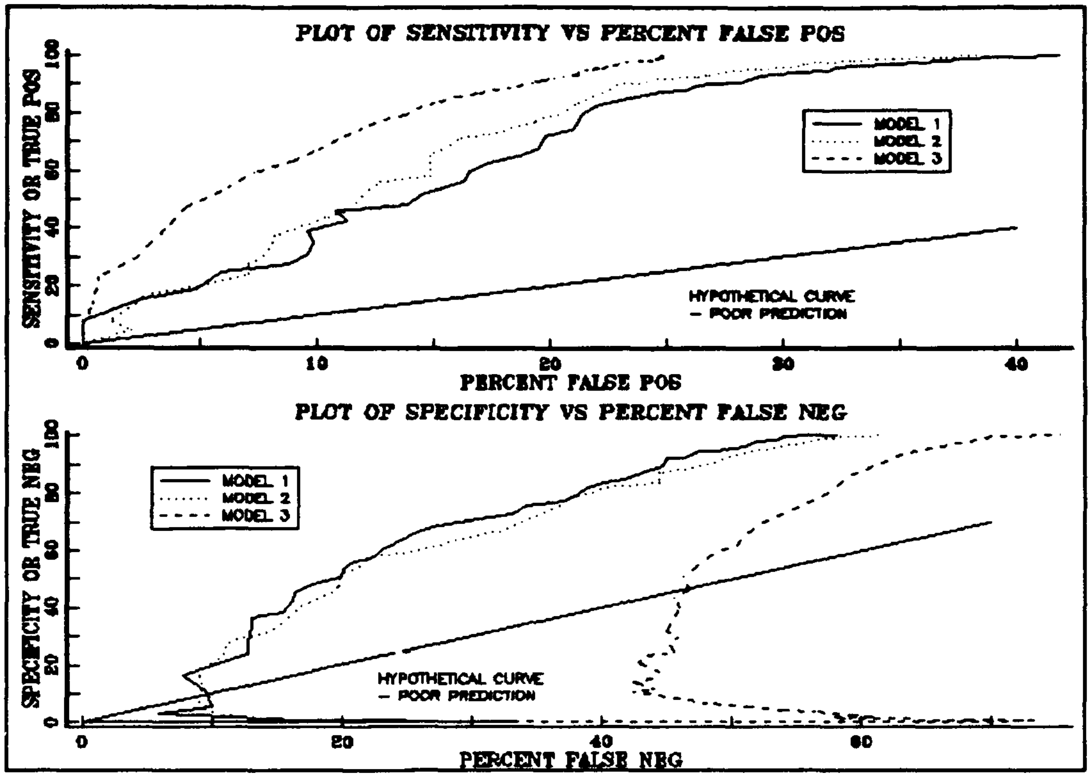

Figure 14: Comparison of Performance Binary Response Models.

APPENDIX C: SAS PROGRAMS AND OUTPUTS

A. BINARY RESPONSE MODEL

1. Models for Current Estimated Potential

```
//LOGREG1 JOB CLASS=A,USER=S6599,PASSWORD=LEE
//*MAIN  LINES=(99)
//  EXEC SAS
//EXTFIN1 DD DISP=SHR,DSN=MSS.S6599.GEN.DATA
//EXTFIN2 DD DISP=SHR,DSN=MSS.S6599.CEP.DATA
//EXTFIN3 DD DISP=SHR,DSN=MSS.S6599.PERF.DATA
//SYSIN DD *

OPTIONS LS=80;
DATA GENREC;
INFILE EXTFIN1;
INPUT
   @1  ID        4.
   @12 DRANK     2.
   @19 DOE       2.
   @36 LEFT      2.
   @43 STATUS    1.
   @45 EDU       1.
   @50 TRAWD     1.
   @58 RANK      1.
   @66 AGE       2.
   @71 SGD       2.
   ;

DATA CEPREC;
INFILE EXTFIN2;
INPUT
   @1  ID        4.
   @6  C92       2.
   @10 C91       2.
   @14 C90       2.
   @18 C89       2.
   ;

DATA PERFREC;
INFILE EXTFIN3;
INPUT
   @1  ID        4.
   @8  P92       2.
   @12 P91       2.
   @16 P90       2.
```

```
    @20  P89            2.
    ;

DATA OFFREC;
    MERGE GENREC CEPREC PERFREC;
    BY ID;

LGSVC = 92 - DOE;
RSNR = 92 - DRANK;
IF (RSNR EQ 92) THEN RSNR = . ;

IF (TRAWD EQ 0) THEN AWARD = 1 ;
IF (TRAWD NE 0) THEN AWARD = 2 ;

DATA ONE; SET OFFREC;
IF (STATUS NE 1) THEN DELETE ;

IF (C92 LT 5) THEN CEP92 = 0 ;
IF (C92 GE 5) THEN CEP92 = 1 ;

TITLE 'BINARY RESPONSE MODEL - CEP MODEL #1' ;
TITLE2 'EVENT=CEP OF MAJ AND BELOW  NON-EVENT=CEP  OF LTC AND ABOVE' ;

PROC LOGISTIC DATA=ONE OUTEST=BETAS1 COVOUT ;
MODEL CEP92  = EDU AWARD RANK LGSVC RSNR AGE SGD
          P91 P90 P89 C91 C90 C89
        / SELECTION=STEPWISE
          SLE=0.1
          SLS=0.12
          DETAILS
          CTABLE ;

PROC PRINT DATA=BETAS1 ;
TITLE2 'PARAMETER ESTIMATES AND COVARIANCE MATRIX - MODEL 1' ;

PROC LOGISTIC DATA=ONE OUTEST=BETAS2 COVOUT ;
TITLE 'BINARY RESPONSE MODEL - CEP MODEL #2' ;
TITLE2 'EVENT=CEP OF MAJ AND BELOW  NON-EVENT=CEP  OF LTC AND ABOVE' ;
MODEL CEP92  = EDU AWARD RANK LGSVC RSNR AGE SGD P91 C91
        / SELECTION=STEPWISE
          SLE=0.1
          SLS=0.12
          DETAILS
          CTABLE ;

PROC PRINT DATA=BETAS2 ;
TITLE2 'PARAMETER ESTIMATES AND COVARIANCE MATRIX - MODEL 2' ;

PROC LOGISTIC DATA=ONE OUTEST=BETAS3 COVOUT ;
TITLE 'BINARY RESPONSE MODEL - CEP MODEL #3' ;
TITLE2 'EVENT=CEP OF MAJ AND BELOW  NON-EVENT=CEP  OF LTC AND ABOVE' ;
```

```
MODEL CEP92 = EDU AWARD RANK LGSVC RSNR AGE SGD
        / SELECTION=STEPWISE
          SLE=0.1
          SLS=0.12
          DETAILS
          CTABLE ;

PROC PRINT DATA=BETAS3 ;
TITLE2 'PARAMETER ESTIMATES AND COVARIANCE MATRIX - MODEL 3' ;
```

2. Outputs for Current Estimated Potential Models

a. Model 1

BINARY RESPONSE MODEL - CEP MODEL
EVENT=CEP OF MAJ AND BELOW; NON-EVENT=CEP OF LTC AND ABOVE

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	1684.415	868.336	.
SC	1689.534	904.168	.
-2 LOG L	1682.415	854.336	828.080 with 6 DF (p=0.0001)
Score	.	.	612.326 with 6 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCEPT	1	10.3187	0.7885	171.2639	0.0001	.	999.000
RANK	1	-1.4990	0.5690	6.9406	0.0084	-0.547076	0.223
SGD	1	0.7856	0.4086	3.6968	0.0545	0.451408	2.194
P91	1	-0.1831	0.0611	8.9726	0.0027	-0.190943	0.833
P90	1	0.1456	0.0582	6.2590	0.0124	0.147330	1.157
C91	1	-1.4689	0.1314	124.9947	0.0001	-1.063768	0.230
C90	1	-0.6452	0.1268	25.8877	0.0001	-0.452197	0.525

Association of Predicted Probabilities and Observed Responses

Concordant	= 92.6%	Somers' D	= 0.855
Discordant	= 7.0%	Gamma	= 0.859
Tied	= 0.4%	Tau-a	= 0.418
(372186 pairs)		c	= 0.928

Residual Chi-Square = 11.8757 with 7 DF (p=0.1047)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
EDU	0.3294	0.5660
AWARD	0.0011	0.9738
LGSVC	0.0082	0.9280
RSNR	0.1090	0.7413
AGE	1.5338	0.2155
P89	0.9744	0.3236
C89	2.3354	0.1265

NOTE: No (additional) variables met the 0.1 significance level for entry into the model.

Summary of Stepwise Procedure

Step	Variable		Number In	Score Chi-square	Wald Chi-Square	Pr > Chi-Square
	Entered	Removed				
1	C91		1	573.0	.	0.0001
2	C90		2	30.3874	.	0.0001
3	C89		3	3.9100	.	0.0480
4	P91		4	4.3282	.	0.0375
5	P90		5	4.3248	.	0.0376
6	RANK		6	5.1085	.	0.0238
7		C89	5	.	2.1647	0.1412
8	SGD		6	3.7489	.	0.0528

Classification Table

Prob Level	Correct: Event	Correct: Non-Event	Incorrect: Event	Incorrect: Non-Event	Percentages: Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.000	522	0	713	0	42.3	100.0	0.0	57.7	0.0
0.020	516	150	563	6	53.9	98.9	21.0	52.2	3.8
0.040	515	199	514	7	57.8	98.7	27.9	50.0	3.4
0.060	515	208	505	7	58.5	98.7	29.2	49.5	3.3
0.080	515	248	465	7	61.8	98.7	34.8	47.4	2.7
0.100	512	292	421	10	65.1	98.1	41.0	45.1	3.3
0.120	495	382	331	27	71.0	94.8	53.6	40.1	6.6
0.140	491	471	242	31	77.9	94.1	66.1	33.0	6.2
0.160	490	497	216	32	79.9	93.9	69.7	30.6	6.0
0.180	488	519	194	34	81.5	93.5	72.8	28.4	6.1
0.200	484	538	175	38	82.8	92.7	75.5	26.6	6.6
0.220	480	551	162	42	83.5	92.0	77.3	25.2	7.1
0.240	478	565	148	44	84.5	91.6	79.2	23.6	7.2
0.260	474	569	144	48	84.5	90.8	79.8	23.3	7.8
0.280	472	575	138	50	84.8	90.4	80.6	22.6	8.0
0.300	470	589	124	52	85.7	90.0	82.6	20.9	8.1
0.320	468	589	124	54	85.6	89.7	82.6	20.9	8.4
0.340	462	604	109	60	86.3	88.5	84.7	19.1	9.0
0.360	461	605	108	61	86.3	88.3	84.9	19.0	9.2
0.380	452	619	94	70	86.7	86.6	86.8	17.2	10.2
0.400	448	619	94	74	86.4	85.8	86.8	17.3	10.7
0.420	447	627	86	75	87.0	85.6	87.9	16.1	10.7
0.440	444	629	84	78	86.9	85.1	88.2	15.9	11.0
0.460	441	631	82	81	86.8	84.5	88.5	15.7	11.4
0.480	435	633	80	87	86.5	83.3	88.8	15.5	12.1
0.500	432	638	75	90	86.6	82.8	89.5	14.8	12.4
0.520	426	640	73	96	86.3	81.6	89.8	14.6	13.0
0.540	414	647	66	108	85.9	79.3	90.7	13.8	14.3
0.560	410	649	64	112	85.7	78.5	91.0	13.5	14.7
0.580	405	656	57	117	85.9	77.6	92.0	12.3	15.1
0.600	403	658	55	119	85.9	77.2	92.3	12.0	15.3
0.620	401	660	53	121	85.9	76.8	92.6	11.7	15.5
0.640	398	662	51	124	85.8	76.2	92.8	11.4	15.8
0.660	382	667	46	140	84.9	73.2	93.5	10.7	17.3
0.680	369	670	43	153	84.1	70.7	94.0	10.4	18.6
0.700	355	675	38	167	83.4	68.0	94.7	9.7	19.8
0.720	350	676	37	172	83.1	67.0	94.8	9.6	20.3
0.740	347	681	32	175	83.2	66.5	95.5	9.4	20.4
0.760	344	683	30	178	83.2	65.9	95.8	8.0	20.7
0.780	340	683	30	182	82.8	65.1	95.8	8.1	21.0
0.800	329	684	29	193	82.0	63.0	95.9	8.1	22.0
0.820	316	686	27	206	81.1	60.5	96.2	7.9	23.1

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.840	304	689	24	218	80.4	58.2	96.6	7.3	24.0
0.860	292	693	20	230	79.8	55.9	97.2	6.4	24.9
0.880	285	693	20	237	79.2	54.6	97.2	6.6	25.5
0.900	261	696	17	261	77.5	50.0	97.6	6.1	27.3
0.920	176	706	7	346	71.4	33.7	99.0	3.8	32.9
0.940	145	709	4	377	69.1	27.8	99.4	2.7	34.7
0.960	47	713	0	475	61.5	9.0	100.0	0.0	40.0
0.980	33	713	0	489	60.4	6.3	100.0	0.0	40.7
1.000	0	713	0	522	57.7	0.0	100.0	0.0	42.3

b. Model 2

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	2252.754	1124.762	.
SC	2258.156	1162.575	.
-2 LOG L	2250.754	1110.762	1139.992 with 6 DF (p=0.0001)
Score	.	.	787.643 with 6 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCEPT	1	6.3961	0.9465	45.6639	0.0001	.	599.514
EDU	1	-0.1947	0.0696	7.8198	0.0052	-0.143909	0.823
RANK	1	-2.2587	0.2893	60.9497	0.0001	-0.834770	0.104
RSNR	1	-0.2089	0.0505	17.1115	0.0001	-0.269972	0.811
AGE	1	0.2162	0.0405	28.4550	0.0001	0.519284	1.241
P91	1	-0.1833	0.0535	11.7581	0.0006	-0.190574	0.833
C91	1	-1.4499	0.1038	195.0239	0.0001	-1.088819	0.235

Association of Predicted Probabilities and Observed Responses

Concordant	= 93.3%	Somers' D	= 0.867
Discordant	= 6.6%	Gamma	= 0.868
Tied	= 0.2%	Tau-a	= 0.428
(662838 pairs)		c	= 0.933

Residual Chi-Square = 2.5633 with 3 DF (p=0.4640)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
AWARD	1.3497	0.2453
LGSVC	1.0153	0.3136
SGD	0.3037	0.5816

NOTE: No (additional) variables met the 0.1 significance level for entry into the model.

Summary of Stepwise Procedure

Step	Variable Entered	Variable Removed	Number In	Score Chi-square	Wald Chi-Square	Pr > Chi-Square
1	C91		1	766.4	.	0.0001
2	RANK		2	32.9832	.	0.0001
3	AWARD		3	19.8557	.	0.0001
4	P91		4	10.0855	.	0.0015
5	AGE		5	7.1768	.	0.0074
6	RSNR		6	15.2178	.	0.0001
7	EDU		7	3.0566	.	0.0804
8		AWARD	6	.	1.3478	0.2457

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.000	726	0	913	0	44.3	100.0	0.0	55.7	.
0.020	720	214	699	6	57.0	99.2	23.4	49.3	2.7
0.040	719	260	653	7	59.7	99.0	28.5	47.6	2.6
0.060	717	297	616	9	61.9	98.8	32.5	46.2	2.9
0.080	717	350	563	9	65.1	98.8	38.3	44.0	2.5

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.100	711	441	472	15	70.3	97.9	48.3	39.9	3.3
0.120	705	505	408	21	73.8	97.1	55.3	36.7	4.0
0.140	696	565	348	30	76.9	95.9	61.9	33.3	5.0
0.160	688	611	302	38	79.3	94.8	66.9	30.5	5.9
0.180	682	640	273	44	80.7	93.9	70.1	28.6	6.4
0.200	681	660	253	45	81.8	93.8	72.3	27.1	6.4
0.220	673	685	228	53	82.9	92.7	75.0	25.3	7.2
0.240	671	693	220	55	83.2	92.4	75.9	24.7	7.4
0.260	668	709	204	58	84.0	92.0	77.7	23.4	7.6
0.280	658	720	193	68	84.1	90.6	78.9	22.7	8.6
0.300	654	753	160	72	85.8	90.1	82.5	19.7	8.7
0.320	653	754	159	73	85.8	89.9	82.6	19.6	8.8
0.340	646	772	141	80	86.5	89.0	84.6	17.9	9.4
0.360	644	777	136	82	86.7	88.7	85.1	17.4	9.5
0.380	639	785	128	87	86.9	88.0	86.0	16.7	10.0
0.400	633	795	118	93	87.1	87.2	87.1	15.7	10.5
0.420	628	801	112	98	87.2	86.5	87.7	15.1	10.9
0.440	623	805	108	103	87.1	85.8	88.2	14.8	11.3
0.460	621	805	108	105	87.0	85.5	88.2	14.8	11.5
0.480	617	808	105	109	86.9	85.0	88.5	14.5	11.9
0.500	610	814	99	116	86.9	84.0	89.2	14.0	12.5
0.520	607	815	98	119	86.8	83.6	89.3	13.9	12.7
0.540	599	822	91	127	86.7	82.5	90.0	13.2	13.4
0.560	597	824	89	129	86.7	82.2	90.3	13.0	13.5
0.580	586	825	88	140	86.1	80.7	90.4	13.1	14.5
0.600	581	829	84	145	86.0	80.0	90.8	12.6	14.9
0.620	571	829	84	155	85.4	78.7	90.8	12.8	15.8
0.640	553	841	72	173	85.1	76.2	92.1	11.5	17.1
0.660	552	841	72	174	85.0	76.0	92.1	11.5	17.1
0.680	539	844	69	187	84.4	74.2	92.4	11.3	18.1
0.700	537	852	61	189	84.7	74.0	93.3	10.2	18.2
0.720	520	858	55	206	84.1	71.6	94.0	9.6	19.4
0.740	514	867	46	212	84.3	70.8	95.0	8.2	19.6
0.760	494	872	41	232	83.3	68.0	95.5	7.7	21.0
0.780	485	873	40	241	82.9	66.8	95.6	7.6	21.6
0.800	465	880	33	261	82.1	64.0	96.4	6.6	22.9
0.820	453	881	32	273	81.4	62.4	96.5	6.6	23.7
0.840	442	886	27	284	81.0	60.9	97.0	5.8	24.3
0.860	416	890	23	310	79.7	57.3	97.5	5.2	25.8
0.880	395	893	20	331	78.6	54.4	97.8	4.8	27.0
0.900	357	895	18	369	76.4	49.2	98.0	4.8	29.2
0.920	313	902	11	413	74.1	43.1	98.8	3.4	31.4
0.940	154	908	5	572	64.8	21.2	99.5	3.1	38.6
0.960	83	911	2	643	60.6	11.4	99.8	2.4	41.4
0.980	58	913	0	668	59.2	8.0	100.0	0.0	42.3
1.000	0	913	0	726	55.7	0.0	100.0	.	44.3

c. *Model 3*

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	3528.592	2028.528	.
SC	3534.470	2057.915	.
-2 LOG L	3526.592	2018.528	1508.064 with 4 DF (p=0.0001)
Score	.	.	1214.662 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCEPT	1	3.0303	0.5149	34.6427	0.0001	.	20.704
EDU	1	-0.4470	0.0486	84.4948	0.0001	-0.303950	0.640
RANK	1	-3.8708	0.1993	377.2430	0.0001	-1.457808	0.021
RSNR	1	-0.3446	0.0365	89.1981	0.0001	-0.447994	0.708
AGE	1	0.2174	0.0275	62.3203	0.0001	0.610420	1.243

Association of Predicted Probabilities and Observed Responses

Concordant	= 90.3%	Somers' D	= 0.808
Discordant	= 9.5%	Gamma	= 0.810
Tied	= 0.3%	Tau-a	= 0.385
(1654052 pairs)		c	= 0.904

Residual Chi-Square = 4.8395 with 3 DF (p=0.1839)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
AWARD	2.0350	0.1537
LGSVC	1.7257	0.1890
SGD	0.2818	0.5955

NOTE: No (additional) variables met the 0.1 significance level for entry into the model.

Summary of Stepwise Procedure

Step	Variable Entered	Variable Removed	Number In	Score Chi-square	Wald Chi-Square	Pr > Chi-Square
1	RANK		1	1040.7	.	0.0001
2	EDU		2	152.8	.	0.0001
3	RSNR		3	31.0777	.	0.0001
4	AGE		4	66.2695	.	0.0001

Classification Table

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.000	1609	0	1028	0	61.0	100.0	0.0	39.0	.
0.020	1606	109	919	3	65.0	99.8	10.6	36.4	2.7
0.040	1604	128	900	5	65.7	99.7	12.5	35.9	3.8
0.060	1603	157	871	6	66.7	99.6	15.3	35.2	3.7
0.080	1601	181	847	8	67.6	99.5	17.6	34.6	4.2
0.100	1599	239	789	10	69.7	99.4	23.2	33.0	4.0
0.120	[illegible]94	300	728	15	71.8	99.1	29.2	31.4	4.8
0.140	1585	361	667	24	73.8	98.5	35.1	29.6	6.2
0.160	1576	418	610	33	75.6	97.9	40.7	27.9	7.3
0.180	1565	490	538	44	77.9	97.3	47.7	25.6	8.2
0.200	1554	531	497	55	79.1	96.6	51.7	24.2	9.4
0.220	1542	559	469	67	79.7	95.8	54.4	23.3	10.7
0.240	1533	592	436	76	80.6	95.3	57.6	22.1	11.4
0.260	1518	611	417	91	80.7	94.3	59.4	21.6	13.0
0.280	1509	629	399	100	81.1	93.8	61.2	20.9	13.7
0.300	1500	644	384	109	81.3	93.2	62.6	20.4	14.5
0.320	1494	660	368	115	81.7	92.9	64.2	19.8	14.8
0.340	1488	679	349	121	82.2	92.5	66.1	19.0	15.1
0.360	1478	699	329	131	82.6	91.9	68.0	18.2	15.8
0.380	1468	712	316	141	82.7	91.2	69.3	17.7	16.5
0.400	1460	720	308	149	82.7	90.7	70.0	17.4	17.1
0.420	1453	740	288	156	83.2	90.3	72.0	16.5	17.4
0.440	1445	761	267	164	83.7	89.8	74.0	15.6	17.7
0.460	1439	777	251	170	84.0	89.4	75.6	14.9	18.0
0.480	1426	797	231	183	84.3	88.6	77.5	13.9	18.7
0.500	1406	810	218	203	84.0	87.4	78.8	13.4	20.0
0.520	1395	821	207	214	84.0	86.7	79.9	12.9	20.7
0.540	1384	833	195	225	84.1	86.0	81.0	12.3	21.3
0.560	1363	846	182	246	83.8	84.7	82.3	11.8	22.5

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.580	1352	857	171	257	83.8	84.0	83.4	11.2	23.1
0.600	1340	865	163	269	83.6	83.3	84.1	10.8	23.7
0.620	1315	880	148	294	83.2	81.7	85.6	10.1	25.0
0.640	1298	887	141	311	82.9	80.7	86.3	9.8	26.0
0.660	1285	895	133	324	82.7	79.9	87.1	9.4	26.6
0.680	1274	907	121	335	82.7	79.2	88.2	8.7	27.0
0.700	1264	909	119	345	82.4	78.6	88.4	8.6	27.5
0.720	1243	914	114	366	81.8	77.3	88.9	8.4	28.6
0.740	1238	918	110	371	81.8	76.9	89.3	8.2	28.8
0.760	1227	922	106	382	81.5	76.3	89.7	8.0	29.3
0.780	1211	928	100	398	81.1	75.3	90.3	7.6	30.0
0.800	1201	929	99	408	80.8	74.6	90.4	7.6	30.5
0.820	1197	930	98	412	80.7	74.4	90.5	7.6	30.7
0.840	1185	931	97	424	80.2	73.6	90.6	7.6	31.3
0.860	1167	933	95	442	79.6	72.5	90.8	7.5	32.1
0.880	1114	948	80	495	78.2	69.2	92.2	6.7	34.3
0.900	971	962	66	638	73.3	60.3	93.6	6.4	39.9
0.920	716	987	41	893	64.6	44.5	96.0	5.4	47.5
0.940	301	1017	11	1308	50.0	18.7	98.9	3.5	56.3
0.960	58	1026	2	1551	41.1	3.6	99.8	3.3	60.2
0.980	5	1027	1	1604	39.1	0.3	99.9	16.7	61.0
1.000	0	1028	0	1609	39.0	0.0	100.0	.	61.0

3. Models for Performance Appraisal

```
//LOGREG2 JOB CLASS=A,USER=S6599,PASSWORD=LEE
//*MAIN LINES=(99)
// EXEC SAS
//EXTFIN1 DD DISP=SHR,DSN=MSS.S6599.GEN.DATA
//EXTFIN2 DD DISP=SHR,DSN=MSS.S6599.CEP.DATA
//EXTFIN3 DD DISP=SHR,DSN=MSS.S6599.PERF.DATA
//SYSIN DD *

OPTIONS LS=80;
DATA GENREC;
INFILE EXTFIN1;
INPUT
   @1  ID       4.
   @12 DRANK    2.
   @19 DOE      2.
   @36 LEFT     2.
   @43 STATUS   1.
   @45 EDU      1.
   @50 TRAWD    1.
   @58 RANK     1.
```

```
    @66  AGE      2.
    @71  SGD      2.
    ;

DATA CEPREC;
INFILE EXTFIN2;
INPUT
    @1   ID       4.
    @6   C92      2.
    @10  C91      2.
    @14  C90      2.
    @18  C89      2.
    ;

DATA PERFREC;
INFILE EXTFIN3;
INPUT
    @1   ID       4.
    @8   P92      2.
    @12  P91      2.
    @16  P90      2.
    @20  P89      2.
    ;

DATA OFFREC;
    MERGE GENREC CEPREC PERFREC;
    BY ID;

LGSVC = 92 - DOE;
RSNR = 92 - DRANK;
IF (RSNR EQ 92) THEN RSNR = . ;

IF (TRAWD EQ 0) THEN AWARD = 1 ;
IF (TRAWD NE 0) THEN AWARD = 2 ;

DATA ONE; SET OFFREC;
IF (STATUS NE 1) THEN DELETE ;

IF (P92 LT 11) THEN PERF92 = 0 ;
IF (P92 GE 11) THEN PERF92 = 1 ;

TITLE 'BINARY RESPONSE MODEL - PERFORMANCE MODEL #1';
TITLE2 'EVENT=GRADE  B MINUS AND BELOW  NON-EVENT=GRADE  B AND ABOVE';

PROC LOGISTIC DATA=ONE OUTEST=BETAS1 COVOUT ;
MODEL PERF92 = EDU AWARD RANK LGSVC RSNR AGE SGD
           P91 P90 P89 C91 C90 C89
          / SELECTION=STEPWISE
           SLE=0.1
           SLS=0.12
           DETAILS
```

```
        CTABLE ;

PROC PRINT DATA=BETAS1 ;
TITLE2 'PARAMETER ESTIMATES AND COVARIANCE MATRIX - MODEL 1' ;

PROC LOGISTIC DATA=ONE OUTEST=BETAS2 COVOUT ;
TITLE 'BINARY RESPONSE MODEL - PERFORMANCE MODEL #2';
TITLE2 'EVENT=GRADE B MINUS AND BELOW  NON-EVENT=GRADE B AND ABOVE';
MODEL PERF92 = EDU AWARD RANK LGSVC RSNR AGE SGD P91 C91
       / SELECTION=STEPWISE
        SLE=0.1
        SLS=0.12
        DETAILS
        CTABLE ;

PROC PRINT DATA=BETAS2 ;
TITLE2 'PARAMETER ESTIMATES AND COVARIANCE MATRIX - MODEL 2' ;

PROC LOGISTIC DATA=ONE OUTEST=BETAS3 COVOUT ;
TITLE 'BINARY RESPONSE MODEL - PERFORMANCE MODEL #3';
TITLE2 'EVENT=GRADE B MINUS AND BELOW  NON-EVENT=GRADE B AND ABOVE';
MODEL PERF92 = EDU AWARD RANK LGSVC RSNR AGE SGD
       / SELECTION=STEPWISE
        SLE=0.1
        SLS=0.12
        DETAILS
        CTABLE ;

PROC PRINT DATA=BETAS3 ;
TITLE2 'PARAMETER ESTIMATES AND COVARIANCE MATRIX - MODEL 3' ;
```

4. Outputs for Performance Appraisal Models

a. Model 1

BINARY RESPONSE MODEL - PERFORMANCE MODEL
EVENT=GRADE B MINUS AND BELOW; NON-EVENT=GRADE B AND ABOVE

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	1680.554	1256.221	.
SC	1685.673	1292.053	.
-2 LOG L	1678.554	1242.221	436.333 with 6 DF (p=0.0001)
Score	.	.	370.328 with 6 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCEPT	1	6.5316	0.5137	161.6679	0.0001	.	686.504
AWARD	1	0.3560	0.1899	3.5162	0.0608	0.082224	1.428
RANK	1	1.3880	0.1555	79.6384	0.0001	0.506578	4.007
RSNR	1	-0.2445	0.0312	61.3598	0.0001	-0.332231	0.783
P91	1	-0.4399	0.0487	81.6485	0.0001	-0.458702	0.644
C91	1	-0.6446	0.1031	39.1117	0.0001	-0.466811	0.525
C89	1	-0.2815	0.0831	11.4646	0.0007	-0.210184	0.755

Association of Predicted Probabilities and Observed Responses

Concordant	= 82.3%	Somers' D	= 0.648
Discordant	= 17.4%	Gamma	= 0.650
Tied	= 0.3%	Tau-a	= 0.316
(371004 pairs)		c	= 0.824

Residual Chi-Square = 4.7942 with 7 DF (p=0.6851)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
EDU	0.2783	0.5978
LGSVC	0.0522	0.8192
AGE	0.0254	0.8734
SGD	0.1182	0.7310
P90	1.3250	0.2497
P89	0.0100	0.9202
C90	1.5263	0.2167

NOTE: No (additional) variables met the 0.1 significance level for entry into the model.

Summary of Stepwise Procedure

Step	Variable Entered	Variable Removed	Number In	Score Chi-square	Wald Chi-Square	Pr > Chi-Square
1	P91		1	205.1	.	0.0001
2	RSNR		2	95.5027	.	0.0001
3	AGE		3	37.6166	.	0.0001
4	C91		4	41.5234	.	0.0001
5	RANK		5	21.8315	.	0.0001
6		AGE	4	.	0.0334	0.8551
7	C89		5	10.0994	.	0.0015
8	AWARD		6	3.5263	.	0.0604

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.020	719	0	516	0	58.2	100.0	0.0	41.8	.
0.040	718	2	514	1	58.3	99.9	0.4	41.7	33.3
0.060	718	5	511	1	58.5	99.9	1.0	41.6	16.7
0.080	718	16	500	1	59.4	99.9	3.1	41.1	5.9
0.100	716	27	489	3	60.2	99.6	5.2	40.6	10.0
0.120	713	57	459	6	62.3	99.2	11.0	39.2	9.5
0.140	712	83	433	7	64.4	99.0	16.1	37.8	7.8
0.160	707	104	412	12	65.7	98.3	20.2	36.8	10.3
0.180	701	123	393	18	66.7	97.5	23.8	35.9	12.8
0.200	698	144	372	21	68.2	97.1	27.9	34.8	12.7
0.220	693	173	343	26	70.1	96.4	33.5	33.1	13.1
0.240	691	187	329	28	71.1	96.1	36.2	32.3	13.0
0.260	683	196	320	36	71.2	95.0	38.0	31.9	15.5
0.280	677	217	299	42	72.4	94.2	42.1	30.6	16.2
0.300	674	231	285	45	73.3	93.7	44.8	29.7	16.3
0.320	669	242	274	50	73.8	93.0	46.9	29.1	17.1
0.340	663	251	265	56	74.0	92.2	48.6	28.6	18.2
0.360	655	258	258	64	73.9	91.1	50.0	28.3	19.9
0.380	650	275	241	69	74.9	90.4	53.3	27.0	20.1
0.400	645	285	231	74	75.3	89.7	55.2	26.4	20.6
0.420	633	296	220	86	75.2	88.0	57.4	25.8	22.5

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.440	626	310	206	93	75.8	87.1	60.1	24.8	23.1
0.460	618	320	196	101	76.0	86.0	62.0	24.1	24.0
0.480	608	333	183	111	76.2	84.6	64.5	23.1	25.0
0.500	596	346	170	123	76.3	82.9	67.1	22.2	26.2
0.520	586	353	163	133	76.0	81.5	68.4	21.8	27.4
0.540	566	363	153	153	75.2	78.7	70.3	21.3	29.7
0.560	551	369	147	168	74.5	76.6	71.5	21.1	31.3
0.580	534	374	142	185	73.5	74.3	72.5	21.0	33.1
0.600	526	382	134	193	73.5	73.2	74.0	20.3	33.6
0.620	517	388	128	202	73.3	71.9	75.2	19.8	34.2
0.640	484	399	117	235	71.5	67.3	77.3	19.5	37.1
0.660	466	409	107	253	70.9	64.8	79.3	18.7	38.2
0.680	451	421	95	268	70.6	62.7	81.6	17.4	38.9
0.700	431	431	85	288	69.8	59.9	83.5	16.5	40.1
0.720	407	437	79	312	68.3	56.6	84.7	16.3	41.7
0.740	372	453	63	347	66.8	51.7	87.8	14.5	43.4
0.760	347	460	56	372	65.3	48.3	89.1	13.9	44.7
0.780	329	476	40	390	65.2	45.8	92.2	10.8	45.0
0.800	306	477	39	413	63.4	42.6	92.4	11.3	46.4
0.820	281	486	30	438	62.1	39.1	94.2	9.6	47.4
0.840	254	488	28	465	60.1	35.3	94.6	9.9	48.8
0.860	225	492	24	494	58.1	31.3	95.3	9.6	50.1
0.880	197	497	19	522	56.2	27.4	96.3	8.8	51.2
0.900	177	505	11	542	55.2	24.6	97.9	5.9	51.8
0.920	135	509	7	584	52.1	18.8	98.6	4.9	53.4
0.940	113	513	3	606	50.7	15.7	99.4	2.6	54.2
0.960	58	516	0	661	46.5	8.1	100.0	0.0	56.2
0.980	23	516	0	696	43.6	3.2	100.0	0.0	57.4
1.000	0	516	0	719	41.8	0.0	100.0	.	58.2

b. Model 2

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	2185.699	1621.567	.
SC	2191.101	1648.576	.
-2 LOG L	2183.699	1611.567	572.132 with 4 DF (p=0.0001)
Score	.	.	487.594 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCEPT	1	7.1631	0.4127	301.2829	0.0001	.	999.000
RANK	1	1.0954	0.1279	73.3601	0.0001	0.404822	2.990
RSNR	1	-0.2833	0.0274	106.9197	0.0001	-0.366112	0.753
P91	1	-0.4378	0.0412	112.8788	0.0001	-0.455195	0.645
C91	1	-0.7922	0.0757	109.5400	0.0001	-0.594913	0.453

Association of Predicted Probabilities and Observed Responses

Concordant	= 82.3%	Somers' D	= 0.653
Discordant	= 17.0%	Gamma	= 0.657
Tied	= 0.6%	Tau-a	= 0.309
(635670 pairs)		c	= 0.826

Residual Chi-Square = 4.0442 with 5 DF (p=0.5431)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
EDU	0.6176	0.4319
AWARD	0.7092	0.3997
LGSVC	0.5795	0.4465
AGE	0.1949	0.6589
SGD	0.0000	0.9990

NOTE: No (additional) variables met the 0.1 significance level for entry into the model.

Summary of Stepwise Procedure

Step	Variable Entered	Variable Removed	Number In	Score Chi-square	Wald Chi-Square	Pr > Chi-Square
1	P91		1	298.9	.	0.0001
2	RSNR		2	115.1	.	0.0001
3	C91		3	50.0724	.	0.0001
4	RANK		4	78.2818	.	0.0001

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.020	1009	0	630	0	61.6	100.0	0.0	38.4	.
0.040	1008	3	627	1	61.7	99.9	0.5	38.3	25.0
0.060	1008	5	625	1	61.8	99.9	0.8	38.3	16.7
0.080	1007	17	613	2	62.5	99.8	2.7	37.8	10.5
0.100	1005	40	590	4	63.8	99.6	6.3	37.0	9.1
0.120	1003	58	572	6	64.7	99.4	9.2	36.3	9.4
0.140	999	100	530	10	67.1	99.0	15.9	34.7	9.1
0.160	997	117	513	12	68.0	98.8	18.6	34.0	9.3
0.180	992	142	488	17	69.2	98.3	22.5	33.0	10.7
0.200	987	175	455	22	70.9	97.8	27.8	31.6	11.2
0.220	979	192	438	30	71.4	97.0	30.5	30.9	13.5
0.240	976	196	434	33	71.5	96.7	31.1	30.8	14.4
0.260	964	234	396	45	73.1	95.5	37.1	29.1	16.1
0.280	961	243	387	48	73.5	95.2	38.6	28.7	16.5
0.300	958	255	375	51	74.0	94.9	40.5	28.1	16.7
0.320	943	279	351	66	74.6	93.5	44.3	27.1	19.1
0.340	941	286	344	68	74.9	93.3	45.4	26.8	19.2
0.360	936	296	334	73	75.2	92.8	47.0	26.3	19.8
0.380	929	325	305	80	76.5	92.1	51.6	24.7	19.8
0.400	925	332	298	84	76.7	91.7	52.7	24.4	20.2
0.420	919	336	294	90	76.6	91.1	53.3	24.2	21.1
0.440	906	362	268	103	77.4	89.8	57.5	22.8	22.2
0.460	898	368	262	111	77.2	89.0	58.4	22.6	23.2
0.480	880	375	255	129	76.6	87.2	59.5	22.5	25.6
0.500	870	383	247	139	76.4	86.2	60.8	22.1	26.6
0.520	844	404	226	165	76.1	83.6	64.1	21.1	29.0
0.540	828	410	220	181	75.5	82.1	65.1	21.0	30.6
0.560	805	422	208	204	74.9	79.8	67.0	20.5	32.6
0.580	780	443	187	229	74.6	77.3	70.3	19.3	34.1
0.600	774	445	185	235	74.4	76.7	70.6	19.3	34.6
0.620	759	456	174	250	74.1	75.2	72.4	18.6	35.4
0.640	745	467	163	264	73.9	73.8	74.1	18.0	36.1
0.660	730	477	153	279	73.6	72.3	75.7	17.3	36.9
0.680	718	491	139	291	73.8	71.2	77.9	16.2	37.2
0.700	682	504	126	327	72.4	67.6	80.0	15.6	39.4
0.720	662	514	116	347	71.8	65.6	81.6	14.9	40.3
0.740	587	527	103	422	68.0	58.2	83.7	14.9	44.5
0.760	571	547	83	438	68.2	56.6	86.8	12.7	44.5
0.780	532	556	74	477	66.4	52.7	88.3	12.2	46.2
0.800	458	572	58	551	62.8	45.4	90.8	11.2	49.1
0.820	441	579	51	568	62.2	43.7	91.9	10.4	49.5
0.840	379	596	34	630	59.5	37.6	94.6	8.2	51.4
0.860	314	603	27	695	55.9	31.1	95.7	7.9	53.5

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.880	276	609	21	733	54.0	27.4	96.7	7.1	54.6
0.900	236	612	18	773	51.7	23.4	97.1	7.1	55.8
0.920	203	619	11	806	50.2	20.1	98.3	5.1	56.6
0.940	174	625	5	835	48.7	17.2	99.2	2.8	57.2
0.960	84	629	1	925	43.5	8.3	99.8	1.2	59.5
0.980	47	629	1	962	41.2	4.7	99.8	2.1	60.5
1.000	0	630	0	1009	38.4	0.0	100.0	.	61.6

c. *Model 3*

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	2958.381	2468.773	.
SC	2964.258	2509.914	.
-2 LOG L	2956.381	2454.773	501.608 with 6 DF (p=0.0001)
Score	.	.	486.441 with 6 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCEPT	1	1.4409	0.5272	7.4710	0.0063	.	4.224
EDU	1	-0.3109	0.0511	37.0796	0.0001	-0.211408	0.733
AWARD	1	0.8515	0.1499	32.2763	0.0001	0.224308	2.343
RANK	1	-1.8529	0.3166	34.2527	0.0001	-0.697837	0.157
RSNR	1	-0.4460	0.0350	162.6577	0.0001	-0.579804	0.640
AGE	1	0.1183	0.0250	22.3293	0.0001	0.332137	1.126
SGD	1	0.3700	0.1917	3.7275	0.0535	0.194989	1.448

Association of Predicted Probabilities and Observed Responses

Concordant	= 79.5%	Somers' D	= 0.595
Discordant	= 20.1%	Gamma	= 0.597
Tied	= 0.4%	Tau-a	= 0.222
(1298210 pairs)		c	= 0.797

Residual Chi-Square = 0.2882 with 1 DF (p=0.5914)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
LGSVC	0.2882	0.5914

Summary of Stepwise Procedure

Step	Variable Entered	Variable Removed	Number In	Score Chi-square	Wald Chi-Square	Pr > Chi-Square
1	RSNR		1	314.1	.	0.0001
2	RANK		2	133.5	.	0.0001
3	EDU		3	30.0064	.	0.0001
4	AWARD		4	20.5901	.	0.0001
5	AGE		5	22.6295	.	0.0001
6	SGD		6	3.7328	.	0.0534

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensitivity	Specificity	False POS	False NEG
0.000	1982	0	655	0	75.2	100.0	0.0	24.8	.
0.020	1981	0	655	1	75.1	99.9	0.0	24.8	100.0
0.040	1978	0	655	4	75.0	99.8	0.0	24.9	100.0
0.060	1978	0	655	4	75.0	99.8	0.0	24.9	100.0
0.080	1976	0	655	6	74.9	99.7	0.0	24.9	100.0
0.100	1975	3	652	7	75.0	99.6	0.5	24.8	70.0
0.120	1971	4	651	11	74.9	99.4	0.6	24.8	73.3
0.140	1969	6	649	13	74.9	99.3	0.9	24.8	68.4
0.160	1969	9	646	13	75.0	99.3	1.4	24.7	59.1
0.180	1967	11	644	15	75.0	99.2	1.7	24.7	57.7
0.200	1963	12	643	19	74.9	99.0	1.8	24.7	61.3
0.220	1956	18	637	26	74.9	98.7	2.7	24.6	59.1
0.240	1951	24	631	31	74.9	98.4	3.7	24.4	56.4
0.260	1950	27	628	32	75.0	98.4	4.1	24.4	54.2
0.280	1945	35	620	37	75.1	98.1	5.3	24.2	51.4
0.300	1942	43	612	40	75.3	98.0	6.6	24.0	48.2
0.320	1936	54	601	46	75.5	97.7	8.2	23.7	46.0

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.340	1932	68	587	50	75.8	97.5	10.4	23.3	42.4
0.360	1923	75	580	59	75.8	97.0	11.5	23.2	44.0
0.380	1914	92	563	68	76.1	96.6	14.0	22.7	42.5
0.400	1904	100	555	78	76.0	96.1	15.3	22.6	43.8
0.420	1889	116	539	93	76.0	95.3	17.7	22.2	44.5
0.440	1880	136	519	102	76.5	94.9	20.8	21.6	42.9
0.460	1865	153	502	117	76.5	94.1	23.4	21.2	43.3
0.480	1849	158	497	133	76.1	93.3	24.1	21.2	45.7
0.500	1836	178	477	146	76.4	92.6	27.2	20.6	45.1
0.520	1818	194	461	164	76.3	91.7	29.6	20.2	45.8
0.540	1810	210	445	172	76.6	91.3	32.1	19.7	45.0
0.560	1790	231	424	192	76.6	90.3	35.3	19.2	45.4
0.580	1759	262	393	223	76.6	88.7	40.0	18.3	46.0
0.600	1745	283	372	237	76.9	88.0	43.2	17.6	45.6
0.620	1717	303	352	265	76.6	86.6	46.3	17.0	46.7
0.640	1696	332	323	286	76.9	85.6	50.7	16.0	46.3
0.660	1652	363	292	330	76.4	83.4	55.4	15.0	47.6
0.680	1618	384	271	364	75.9	81.6	58.6	14.3	48.7
0.700	1571	403	252	411	74.9	79.3	61.5	13.8	50.5
0.720	1537	427	228	445	74.5	77.5	65.2	12.9	51.0
0.740	1485	454	201	497	73.5	74.9	69.3	11.9	52.3
0.760	1426	476	179	556	72.1	71.9	72.7	11.2	53.9
0.780	1332	505	150	650	69.7	67.2	77.1	10.1	56.3
0.800	1254	532	123	728	67.7	63.3	81.2	8.9	57.8
0.820	1173	562	93	809	65.8	59.2	85.8	7.3	59.0
0.840	1067	585	70	915	62.6	53.8	89.3	6.2	61.0
0.860	934	612	43	1048	58.6	47.1	93.4	4.4	63.1
0.880	807	624	31	1175	54.3	40.7	95.3	3.7	65.3
0.900	593	641	14	1389	46.8	29.9	97.9	2.3	68.4
0.920	458	652	3	1524	42.1	23.1	99.5	0.7	70.0
0.940	278	654	1	1704	35.3	14.0	99.8	0.4	72.3
0.960	136	655	0	1846	30.0	6.9	100.0	0.0	73.8
0.980	1	655	0	1981	24.9	0.1	100.0	0.0	75.2
1.000	0	655	0	1982	24.8	0.0	100.0	.	75.2

B. POLYTOMOUS RESPONSE MODEL

1. Current Estimated Potential Model

a. Predicted Probabilities and 95% Confidence Intervals

```
//LOGREG3 JOB CLASS=A,USER=S6599,PASSWORD=LEE
//*MAIN LINES=(99)
// EXEC SAS
//EXTFIN1 DD DISP=SHR,DSN=MSS.S6599.GEN.DATA
```

```
//EXTFIN2 DD DISP=SHR,DSN=MSS.S6599.CEP.DATA
//EXTFIN3 DD DISP=SHR,DSN=MSS.S6599.PERF.DATA
//SYSIN DD *

OPTIONS LS=80;
DATA GENREC;
INFILE EXTFIN1;
INPUT
    @1  ID        4.
    @12 DRANK     2.
    @19 DOE       2.
    @36 LEFT      2.
    @43 STATUS    1.
    @45 EDU       1.
    @50 TRAWD     1.
    @58 RANK      1.
    @66 AGE       2.
    @71 SGD       2.
    ;

DATA CEPREC;
INFILE EXTFIN2;
INPUT
    @1  ID        4.
    @6  C92       2.
    @10 C91       2.
    @14 C90       2.
    @18 C89       2.
    ;

DATA PERFREC;
INFILE EXTFIN3;
INPUT
    @1  ID        4.
    @8  P92       2.
    @12 P91       2.
    @16 P90       2.
    @20 P89       2.
    ;

DATA OFFREC;
    MERGE GENREC CEPREC PERFREC;
    BY ID;

IF (STATUS NE 1) THEN DELETE ;

LGSVC = 92 - DOE;
RSNR = 92 - DRANK;
IF (RSNR EQ 92) THEN RSNR = . ;

IF (TRAWD EQ 0) THEN AWARD = 1 ;
```

```
IF (TRAWD NE 0) THEN AWARD = 2 ;

DATA ONE (DROP = LEFT STATUS) TWO; SET OFFREC;


IF (C92 LE 2) THEN CEP92 = 1 ;
IF (C92 GT 2 AND C92 LE 4) THEN CEP92 = 2 ;
IF (C92 GT 4 AND C92 LE 6) THEN CEP92 = 3 ;
IF (C92 GT 6) THEN CEP92 = 4 ;

IF RANUNI(12345678) LT 0.45 THEN OUTPUT ONE;
ELSE OUTPUT TWO;

TITLE 'STEPWISE LOGISTIC REGRESSION - POLYTOMOUS RESPONSE MODEL';
TITLE2 'CURRENT ESTIMATED POTENTIAL MODEL' ;
TITLE3 '1=CPT RANK  2=MAJ RANK  3=LTC RANK  4=COL AND ABOVE RANK' ;

PROC LOGISTIC DATA=ONE OUTEST=BETAS COVOUT ;
MODEL CEP92 = EDU AWARD RANK LGSVC RSNR AGE SGD P91 C91
        / SELECTION=STEPWISE
         SLE=0.1
         SLS=0.12
         DETAILS
         ;

OUTPUT OUT=PRED P=PHAT LOWER=LCL UPPER=UCL ;

PROC PRINT DATA=BETAS ;
TITLE3 'PARAMETER ESTIMATES AND COVARIANCE MATRIX' ;

PROC PRINT DATA=PRED ;
TITLE3 'PREDICTED PROBABILITIES AND 95% CONFIDENCE LIMITS' ;
```

b. Verification Program

```
//VERIFY3 JOB CLASS=A,USER=S6599,PASSWORD=LEE
//*MAIN LINES=(99)
// EXEC SAS
//EXTFIN1 DD DISP=SHR,DSN=MSS.S6599.GEN.DATA
//EXTFIN2 DD DISP=SHR,DSN=MSS.S6599.CEP.DATA
//EXTFIN3 DD DISP=SHR,DSN=MSS.S6599.PERF.DATA
//SYSIN DD *

OPTIONS LS=80;
DATA GENREC;
INFILE EXTFIN1;
INPUT
   @1  ID        4.
   @12 DRANK     2.
   @19 DOE       2.
```

```
    @36  LEFT         2.
    @43  STATUS       1.
    @45  EDU          1.
    @50  TRAWD        1.
    @58  RANK         1.
    @66  AGE          2.
    @71  SGD          2.
    ;

DATA CEPREC;
INFILE EXTFIN2;
INPUT
    @1   ID           4.
    @6   C92          2.
    @10  C91          2.
    @14  C90          2.
    @18  C89          2.
    ;

DATA PERFREC;
INFILE EXTFIN3;
INPUT
    @1   ID           4.
    @8   P92          2.
    @12  P91          2.
    @16  P90          2.
    @20  P89          2.
    ;

DATA OFFREC;
    MERGE GENREC CEPREC PERFREC;
    BY ID;

IF (STATUS NE 1) THEN DELETE ;

LGSVC = 92 - DOE;
RSNR = 92 - DRANK;
IF (RSNR EQ 92) THEN RSNR = . ;

IF (TRAWD EQ 0) THEN AWARD = 1 ;
IF (TRAWD NE 0) THEN AWARD = 2 ;

DATA ONE TWO; SET OFFREC;

IF (C92 LE 2) THEN CEP92 = 1 ;
IF (C92 GT 2 AND C92 LE 4) THEN CEP92 = 2 ;
IF (C92 GT 4 AND C92 LE 6) THEN CEP92 = 3 ;
IF (C92 GT 6) THEN CEP92 = 4 ;

IF RANUNI(12345678) LT 0.45 THEN OUTPUT ONE;
ELSE OUTPUT TWO;
```

```
DATA THREE; SET ONE;

        INT1 = -0.2548;
        INT2 =  4.1478;
        INT3 =  9.7599;

        BTX = EDU * (-0.1952) + RANK * (-2.2155) + RSNR * (-0.1560)
            + AGE * (0.2379) + P91 * (-0.1374) + C91 * (-1.2298)  ;

NUM1 = EXP(INT1+BTX) ;
DEN1 = (1+NUM1) ;
GAMMA1 = NUM1/DEN1 ;

NUM2 = EXP(INT2+BTX) ;
DEN2 = (1+NUM2) ;
GAMMA2 = NUM2/DEN2 ;

NUM3 = EXP(INT3+BTX) ;
DEN3 = (1+NUM3) ;
GAMMA3 = NUM3/DEN3 ;

P1 = GAMMA1 ;
P2 = GAMMA2 - GAMMA1 ;
P3 = GAMMA3 - GAMMA2 ;
P4 = 1 - GAMMA3 ;

DATA FOUR (KEEP ID CEP92 GAMMA1 GAMMA2 GAMMA3 P1 P2 P3 P4); SET THREE;
PROC PRINT;
```

c. *Cross-Validation of Model*

```
//XVALID3 JOB CLASS=A,USER=S6599,PASSWORD=LEE
//*MAIN  LINES=(99)
//  EXEC SAS
//EXTFIN1 DD DISP=SHR,DSN=MSS.S6599.GEN.DATA
//EXTFIN2 DD DISP=SHR,DSN=MSS.S6599.CEP.DATA
//EXTFIN3 DD DISP=SHR,DSN=MSS.S6599.PERF.DATA
//SYSIN DD *

OPTIONS LS=80;
DATA GENREC;
INFILE EXTFIN1;
INPUT
   @1  ID        4.
   @12 DRANK     2.
   @19 DOE       2.
   @36 LEFT      2.
   @43 STATUS    1.
   @45 EDU       1.
   @50 TRAWD     1.
```

```
    @58  RANK       1.
    @66  AGE        2.
    @71  SGD        2.
    ;

DATA CEPREC;
INFILE EXTFIN2;
INPUT
    @1   ID         4.
    @6   C92        2.
    @10  C91        2.
    @14  C90        2.
    @18  C89        2.
    ;

DATA PERFREC;
INFILE EXTFIN3;
INPUT
    @1   ID         4.
    @8   P92        2.
    @12  P91        2.
    @16  P90        2.
    @20  P89        2.
    ;

DATA OFFREC;
    MERGE GENREC CEPREC PERFREC;
    BY ID;

IF (STATUS NE 1) THEN DELETE ;

LGSVC = 92 - DOE;
RSNR = 92 - DRANK;
IF (RSNR EQ 92) THEN RSNR = . ;

IF (TRAWD EQ 0) THEN AWARD = 1 ;
IF (TRAWD NE 0) THEN AWARD = 2 ;

DATA ONE TWO; SET OFFREC;

IF (C92 LE 2) THEN CEP92 = 1 ;
IF (C92 GT 2 AND C92 LE 4) THEN CEP92 = 2 ;
IF (C92 GT 4 AND C92 LE 6) THEN CEP92 = 3 ;
IF (C92 GT 6) THEN CEP92 = 4 ;

IF RANUNI(12345678) LT 0.45 THEN OUTPUT ONE;
ELSE OUTPUT TWO;

DATA THREE; SET TWO;

        INT1 = -0.2548;
```

```
        INT2 =  4.1478;
        INT3 =  9.7599;

        BTX = EDU * (-0.1952) + RANK * (-2.2155) + RSNR * (-0.1560)
            + AGE * (0.2379) + P91 * (-0.1374) + C91 * (-1.2298)  ;

NUM1 = EXP(INT1+BTX)  ;
DEN1 = (1+NUM1) ;
GAMMA1 = NUM1/DEN1  ;

NUM2 = EXP(INT2+BTX)  ;
DEN2 = (1+NUM2) ;
GAMMA2 = NUM2/DEN2 ;

NUM3 = EXP(INT3+BTX)  ;
DEN3 = (1+NUM3) ;
GAMMA3 = NUM3/DEN3  ;

P1 = GAMMA1  ;
P2 = GAMMA2  - GAMMA1  ;
P3 = GAMMA3  - GAMMA2  ;
P4 = 1 - GAMMA3  ;

IF (P1 EQ .) THEN GROUP = . ;
ELSE IF (P1 GT P2) AND (P1 GT P3) AND (P1 GT P4) THEN GROUP = 1 ;
ELSE IF (P2 GT P1) AND (P2 GT P3) AND (P2 GT P4) THEN GROUP = 2 ;
ELSE IF (P3 GT P1) AND (P3 GT P2) AND (P3 GT P4) THEN GROUP = 3 ;
ELSE GROUP = 4 ;

IF (GROUP EQ .) THEN MATCH = 'MISSING' ;
ELSE IF CEP92 EQ GROUP THEN MATCH = 'CORRECT' ;
ELSE MATCH = 'WRONG' ;
DATA FOUR (KEEP ID CEP92 P1 P2 P3 P4 GROUP MATCH); SET THREE;
PROC PRINT;

TITLE 'ONE WAY FREQUENCY TABLE' ;
PROC FREQ ;
        TABLES MATCH ;
RUN;
```

2. PERFORMANCE MODEL

a. Predicted Probabilities and 95% Confidence Intervals

```
//LOGREG4 JOB CLASS=A,USER=S6599,PASSWORD=LEE
//*MAIN  LINES=(99)
//  EXEC SAS
//EXTFIN1 DD DISP=SHR,DSN=MSS.S6599.GEN.DATA
//EXTFIN2 DD DISP=SHR,DSN=MSS.S6599.CEP.DATA
//EXTFIN3 DD DISP=SHR,DSN=MSS.S6599.PERF.DATA
```

```
//SYSIN DD *

OPTIONS LS=80;
DATA GENREC;
INFILE EXTFIN1;
INPUT
   @1   ID        4.
   @12  DRANK     2.
   @19  DOE       2.
   @36  LEFT      2.
   @43  STATUS    1.
   @45  EDU       1.
   @50  TRAWD     1.
   @58  RANK      1.
   @66  AGE       2.
   @71  SGD       2.
   ;

DATA CEPREC;
INFILE EXTFIN2;
INPUT
   @1   ID        4.
   @6   C92       2.
   @10  C91       2.
   @14  C90       2.
   @18  C89       2.
   ;

DATA PERFREC;
INFILE EXTFIN3;
INPUT
   @1   ID        4.
   @8   P92       2.
   @12  P91       2.
   @16  P90       2.
   @20  P89       2.
   ;

DATA OFFREC;
   MERGE GENREC CEPREC PERFREC;
   BY ID;

IF (STATUS NE 1) THEN DELETE ;

LGSVC = 92 - DOE;
RSNR = 92 - DRANK;
IF (RSNR EQ 92) THEN RSNR = . ;

IF (TRAWD EQ 0) THEN AWARD = 1 ;
IF (TRAWD NE 0) THEN AWARD = 2 ;
```

```
DATA ONE (DROP = LEFT STATUS) TWO; SET OFFREC;

IF (P92 LE 3) THEN PERF92 = 1 ;
IF (P92 GT 3 AND P92 LE 6) THEN PERF92 = 2 ;
IF (P92 GT 6 AND P92 LE 9) THEN PERF92 = 3 ;
IF (P92 GT 9 AND P92 LE 12) THEN PERF92 = 4 ;
IF (P92 GT 12) THEN PERF92 = 5 ;

IF RANUNI(12345678) LT 0.45 THEN OUTPUT ONE;
ELSE OUTPUT TWO;

TITLE 'STEPWISE LOGISTIC REGRESSION - POLYTOMOUS RESPONSE MODEL';
TITLE2 'PERFORMANCE CLASSIFICATION MODEL' ;
TITLE3 '1=E GRADE  2=D GRADE  3=C GRADE  4=B GRADE  5=A GRADE' ;

PROC LOGISTIC DATA=ONE OUTEST=BETAS COVOUT ;
MODEL PERF92 = EDU AWARD RANK LGSVC RSNR AGE SGD P91 C91
        / SELECTION=STEPWISE
          SLE=0.1
          SLS=0.12
          DETAILS
          ;

OUTPUT OUT=PRED P=PHAT LOWER=LCL UPPER=UCL ;

PROC PRINT DATA=BETAS ;
TITLE3 'PARAMETER ESTIMATES AND COVARIANCE MATRIX' ;

PROC PRINT DATA=PRED ;
TITLE3 'PREDICTED PROBABILITIES AND 95% CONFIDENCE LIMITS' ;
```

b. Verification Program

```
//VERIFY4 JOB CLASS=A,USER=S6599,PASSWORD=LEE
//*MAIN  LINES=(99)
//  EXEC SAS
//EXTFIN1 DD DISP=SHR,DSN=MSS.S6599.GEN.DATA
//EXTFIN2 DD DISP=SHR,DSN=MSS.S6599.CEP.DATA
//EXTFIN3 DD DISP=SHR,DSN=MSS.S6599.PERF.DATA
//SYSIN DD *

OPTIONS LS=80;
DATA GENREC;
INFILE EXTFIN1;
INPUT
   @1  ID        4.
   @12 DRANK     2.
   @19 DOE       2.
   @36 LEFT      2.
   @43 STATUS    1.
```

```
    @45  EDU        1.
    @50  TRAWD      1.
    @58  RANK       1.
    @66  AGE        2.
    @71  SGD        2.
    ;

DATA CEPREC;
INFILE EXTFIN2;
INPUT
    @1   ID         4.
    @6   C92        2.
    @10  C91        2.
    @14  C90        2.
    @18  C89        2.
    ;

DATA PERFREC;
INFILE EXTFIN3;
INPUT
    @1   ID         4.
    @8   P92        2.
    @12  P91        2.
    @16  P90        2.
    @20  P89        2.
    ;

DATA OFFREC;
    MERGE GENREC CEPREC PERFREC;
    BY ID;

IF (STATUS NE 1) THEN DELETE ;

LGSVC = 92 - DOE;
RSNR = 92 - DRANK;
IF (RSNR EQ 92) THEN RSNR = . ;

IF (TRAWD EQ 0) THEN AWARD = 1 ;
IF (TRAWD NE 0) THEN AWARD = 2 ;

DATA ONE TWO; SET OFFREC;

IF (P92 LE 3) THEN PERF92 = 1 ;
IF (P92 GT 3 AND P92 LE 6) THEN PERF92 = 2 ;
IF (P92 GT 6 AND P92 LE 9) THEN PERF92 = 3 ;
IF (P92 GT 9 AND P92 LE 12) THEN PERF92 = 4 ;
IF (P92 GT 12) THEN PERF92 = 5 ;

IF RANUNI(12345678) LT 0.45 THEN OUTPUT ONE;
ELSE OUTPUT TWO;
```

```
DATA THREE; SET ONE;

        INT1 = 1.2353;
        INT2 = 2.0194;
        INT3 = 5.7756;
        INT4 = 9.8619;
        BTX = RANK*(0.8447) + RSNR*(-0.2118) + P91*(-0.3637) + C91*(-0.5939) ;

NUM1 = EXP(INT1+BTX) ;
DEN1 = (1+NUM1) :
GAMMA1 = NUM1/DEN1 ;

NUM2 = EXP(INT2+BTX) ;
DEN2 = (1+NUM2) ;
GAMMA2 = NUM2/DEN2 ;

NUM3 = EXP(INT3+BTX) ;
DEN3 = (1+NUM3) ;
GAMMA3 = NUM3/DEN3 ;

NUM4 = EXP(INT4+BTX) ;
DEN4 = (1+NUM4) ;
GAMMA4 = NUM4/DEN4 ;

P1 = GAMMA1 ;
P2 = GAMMA2 - GAMMA1 ;
P3 = GAMMA3 - GAMMA2 ;
P4 = GAMMA4 - GAMMA3 ;
P5 = 1 - GAMMA4 ;

DATA FOUR (KEEP ID PERF92 GAMMA1 GAMMA2 GAMMA3 GAMMA4 P1 P2 P3 P4 P5);
SET THREE;
PROC PRINT;
```

c. *Cross-Validation of Model*

```
//XVALID4 JOB CLASS=A,USER=S6599,PASSWORD=LEE
//*MAIN  LINES=(99)
//  EXEC SAS
//EXTFIN1 DD DISP=SHR,DSN=MSS.S6599.GEN.DATA
//EXTFIN2 DD DISP=SHR,DSN=MSS.S6599.CEP.DATA
//EXTFIN3 DD DISP=SHR,DSN=MSS.S6599.PERF.DATA
//SYSIN DD *

OPTIONS LS=80;
DATA GENREC;
INFILE EXTFIN1;
INPUT
   @1  ID          4.
   @12 DRANK       2.
```

```
    @19  DOE        2.
    @36  LEFT       2.
    @43  STATUS     1.
    @45  EDU        1.
    @50  TRAWD      1.
    @58  RANK       1.
    @66  AGE        2.
    @71  SGD        2. ;

DATA CEPREC;
INFILE EXTFIN2;
INPUT
    @1   ID         4.
    @6   C92        2.
    @10  C91        2.
    @14  C90        2.
    @18  C89        2. ;

DATA PERFREC;
INFILE EXTFIN3;
INPUT
    @1   ID         4.
    @8   P92        2.
    @12  P91        2.
    @16  P90        2.
    @20  P89        2. ;

DATA OFFREC;
    MERGE GENREC CEPREC PERFREC;
    BY ID;

IF (STATUS NE 1) THEN DELETE ;

LGSVC = 92 - DOE;
RSNR = 92 - DRANK;
IF (RSNR EQ 92) THEN RSNR = . ;

IF (TRAWD EQ 0) THEN AWARD = 1 ;
IF (TRAWD NE 0) THEN AWARD = 2 ;

DATA ONE TWO; SET OFFREC;

IF (P92 LE 3) THEN PERF92 = 1 ;
IF (P92 GT 3 AND P92 LE 6) THEN PERF92 = 2 ;
IF (P92 GT 6 AND P92 LE 9) THEN PERF92 = 3 ;
IF (P92 GT 9 AND P92 LE 12) THEN PERF92 = 4 ;
IF (P92 GT 12) THEN PERF92 = 5 ;

IF RANUNI(12345678) LT 0.45 THEN OUTPUT ONE;
ELSE OUTPUT TWO;
```

```
DATA THREE; SET TWO;

        INT1 = 1.2353;
        INT2 = 2.0194;
        INT3 = 5.7756;
        INT4 = 9.8619;
        BTX = RANK*(0.8447) + RSNR*(-0.2118) + P91*(-0.3637) + C91*(-0.5939) ;

NUM1 = EXP(INT1+BTX) ;
DEN1 = (1+NUM1) ;
GAMMA1 = NUM1/DEN1 ;

NUM2 = EXP(INT2+BTX) ;
DEN2 = (1+NUM2) ;
GAMMA2 = NUM2/DEN2 ;

NUM3 = EXP(INT3+BTX) ;
DEN3 = (1+NUM3) ;
GAMMA3 = NUM3/DEN3 ;

NUM4 = EXP(INT4+BTX) ;
DEN4 = (1+NUM4) ;
GAMMA4 = NUM4/DEN4 ;

P1 = GAMMA1 ;
P2 = GAMMA2 - GAMMA1 ;
P3 = GAMMA3 - GAMMA2 ;
P4 = GAMMA4 - GAMMA3 ;
P5 = 1 - GAMMA4 ;

IF (P1 EQ .) THEN GROUP = . ;
ELSE IF (P1 GT P2) AND (P1 GT P3) AND (P1 GT P4) AND (P1 GT P5)
                                    THEN GROUP = 1 ;
ELSE IF (P2 GT P1) AND (P2 GT P3) AND (P2 GT P4) AND (P2 GT P5)
                                    THEN GROUP = 2 ;
ELSE IF (P3 GT P1) AND (P3 GT P2) AND (P3 GT P4) AND (P3 GT P5)
                                    THEN GROUP = 3 ;
ELSE IF (P4 GT P1) AND (P4 GT P2) AND (P4 GT P3) AND (P4 GT P5)
                                    THEN GROUP = 4 ;
ELSE GROUP = 5 ;

IF (GROUP EQ .) THEN MATCH = 'MISSING' ;
ELSE IF PERF92 EQ GROUP THEN MATCH = 'CORRECT' ;
ELSE MATCH = 'WRONG' ;
DATA FOUR (KEEP ID PERF92 P1 P2 P3 P4 P5 GROUP MATCH); SET THREE;
PROC PRINT;

TITLE 'ONE WAY FREQUENCY TABLE' ;
PROC FREQ ;
        TABLES MATCH ;
RUN;
```

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria VA 22304-6145	2
2.	Library, Code 052 Naval Postgraduate School Monterey CA 93943-5002	2
3.	Chief Defence Scientist Mindef Bldg, Gombak Drive Off Upper Bukit Timah Road Singapore 2366	1
4.	Head Operations Analysis Department Mindef Bldg, Gombak Drive Off Upper Bukit Timah Road Singapore 2366	2
5.	Head Navy Operations Analysis Branch Mindef Bldg, Gombak Drive Off Upper Bukit Timah Road Singapore 2366	2
6.	Chief of Navy Mindef Bldg, Gombak Drive Off Upper Bukit Timah Road Singapore 2366	1
7.	Director Manpower Mindef Bldg, Gombak Drive Off Upper Bukit Timah Road Singapore 2366	1
8.	Professor So Young Sohn (Code OR/Sh) Operations Analysis Dept. Naval Postgraduate School Monterey, CA 93943	2
9.	Professor G. Lindsay (Code OR/Ls) Operations Analysis Dept. Naval Postgraduate School Monterey, CA 93943	2